

Is science a job for governments?

The British government's call for advice on science policy raises the question of what government's role should be.

THE British government is not alone in seeking to enhance prosperity by the more effective application of science, research in particular. President Bill Clinton has spoken more often of "technology transfer" than of research. Among European governments, the French has been the most successful; since President François Mitterrand's election in 1980, strong links have been forged between research institutes and industrial companies while revitalizing much of academic research — but at considerable cost in annual budgets. The experience of Japan, on the other hand, impressive and even startling though it is, is less relevant. Even in recent decades, Japan's technology has rested not on a transfer but a recruitment policy. Bright young people have been recruited into lifelong careers of training and hard work by companies persuaded that only technology could allow them to succeed. Only now has it become apparent that the policy is too shallow for the long run, and that the government must invest in basic science.

The British determination to reorganize the part of the scientific enterprise the government controls directly is nevertheless warmly to be welcomed. Taxpayers will rejoice if the edge can be taken off seemingly chronic economic difficulties. The scientific community would also like to be rid of the common suspicion that it consists of a bunch of layabouts, unaccountable spenders of public funds. But there are dangers. The article on page 581, which is meant to provide background reading for those attending public meetings in Edinburgh this week and London next month, draws attention to some of the risks. Nobody denies that the impending reorganization is important. It could bring huge benefits but, if misconceived, it could do great damage. Is it not always thus with important matters of public policy?

The chief danger is that the two partners in any scientific enterprise — basic research and its application — will be dealt with as if they are at opposite poles of a wide spectrum. The ground for that suspicion is the frequently expressed opinion of British industrialists, openly reflected in some of the advice the government has been given, that basic research, being untied to immediate goals, is irrelevant to wealth-creation. That calculation overlooks not just the importance of basic science as a source of innovation of technique (which can in principle be picked up from books and journals) but the role of basic research in the training of able young men and women.

That is why this journal has consistently held that one ingredient of a new policy for British research should be to make the training of people the centrepiece of the justification of basic research. There are awkward implications, of course. Graduate students' stipends would have to be increased. (Why should the people to whom Britain trusts its future be paid less than a half

of the average wage?) Graduate students would have to qualify for their right to embark on a research degree, but then would have more freedom than at present to influence their own course of studies. It would also make sense that public support for basic science should be partly redirected towards those fields of research in which the gap between discovery and application is smallest, not so much for the innovations that would flow but for the increased supply of skill that there would be. (In molecular biology now, the gap seems to have vanished altogether.)

But it would be folly to suppose that reforms of this kind, long overdue, would be enough to win the British government the prize it seeks. It is all too well known that most British companies have yet to be persuaded of the truths their Japanese peers live by in what is now the global economy — that the products that command attention are those developed with daring and almost endless trouble. Britain, by contrast, has a long history of half-baked development, from the Comet aircraft of the 1950s to the airborne radar system abandoned three years ago in favour of the superior US AWACS — both of these in fields in which the British government has lavished substantial funds. One of the government's problems for the weeks ahead is to find ways of making British companies pay attention to the importance of research and development.

A greater difficulty lies closer to the government's own doorstep. Inevitably, the administration of government policy falls on the shoulder of officials. And while the British civil service is proud of its incorruptible intelligence, it is far from being technically literate. Too often, the people whose duty is to advise on the fate of great projects do not properly understand what they are about. That fault, coupled with the habit of secretiveness that still pervades the British habit of government, means that many decisions are ill-judged. For the same reasons, the government is less than able to take the initiative on technical matters. By good fortune the minister in charge of science also has responsibility for the British civil service. Something must be done about it.

But is it not the proper role of government to create the climate in which science and technology can prosper, and then to stand aside? That is the present government's belief. But in science as in other fields (monetary policy, for example) the past ten years have not created the conditions under which the government can let problems solve themselves. Cajoling industry into taking research and development seriously will require that money should be spent, preferably on projects rather than tax-breaks. Outside industries such as pharmaceuticals, British industry also needs some illustrations to demonstrate that it can succeed in radical innovations. Talk from on high can help, but it must be perceptive talk, not empty exhortation. □



Britain urged to lift barriers to investment in biotechnology

London. Britain's biotechnology companies plan to ask the government to make it easier for small start-up ventures registered in Britain to attract capital. The effort, which is being coordinated by the BioIndustry Association (BIA), follows growing evidence that difficulties in raising venture capital domestically is forcing small biotechnology companies into the arms of foreign investors, particularly those from the United States.

Two weeks ago, for example, the Aberdeen-based company Scotgen, which has developed a range of products using "humanized" antibodies that derive from research financed by the Medical Research Council, announced that it was merging with the biotechnology company Vasocor, Inc. of Menlo Park, California. The new company will be called Scotgen Biopharmaceuticals, Inc.

The merger has already allowed the new company to raise an additional \$7 million in capital, from US investors. Most of the money will go towards clinical trials.

The injection of capital is a windfall for Scotgen's original investors. The University of Aberdeen, which provided laboratory facilities and other assistance in return for an equity stake, could see its investment of tens of thousands of pounds grow in value to several million. Similar gains could accrue to the Scottish Development Agency, another early backer.

But the merger and refinancing also means that US investors will dominate the company's new board of directors. Although Bill Harris, professor of genetics at the University of Aberdeen and founder of Scotgen, who will act as both president and chief scientific officer of the merged company, regrets this shift, he says that "we did not even look in the UK" for new financing because of a previous lack of interest.

The merge also gives Scotgen a US presence that will help it at some point to obtain a public listing on the National Association of Securities Dealers Quotations (NASDAQ) exchange. Cantab Pharmaceuticals of Cambridge pursued the same route last summer when it went public in the United States.

A US listing gives a company greater access to a more aggressive pool of investors and is easier to obtain than a listing on the British Stock Exchange, which places stricter demands on fledgling companies. One particularly difficult standard is the

need to show a record of profit making. A recent report by the Advisory Committee on Science and Technology (ACOST) suggesting changes in Stock Exchange rules drew attention to the lack of an efficient "exit

change proposed new guidelines that would, for example, drop the profitable trading requirement. But the exchange still wants companies that apply to hold (or expect to gain) the rights to a number of patents and to have at least two drugs in clinical trials.

The comment period on these proposals ended last Friday and the response has been mixed. Several companies (and the BIA) have welcomed their general thrust, and have proposed some minor changes. But financial analysts are concerned that the rules are likely to remain relatively conservative.

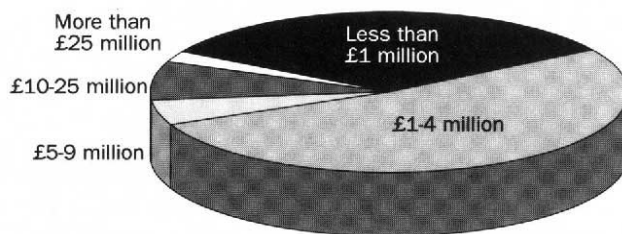
"If the situation continues, the City of London will continue to lose out on a great opportunity", says Caroline Vaughan of Newmarket Management Services. "We must be among the top two or three countries in the world in terms of technical capability [in biotechnology], but the downstream benefit is going to go to NASDAQ and the US." The director of the BIA, Louis da Gama, says that the BIA intends to ask the Department of Trade and Industry to address a range of issues that make investment opportunities in biotechnology less attractive in Britain.

In the meantime, scientists working in university and research council laboratories report an increasing number of enquiries from foreign investors. In several cases, the offers would involve buying the rights to fundamental research results and developing them in US-based laboratories.

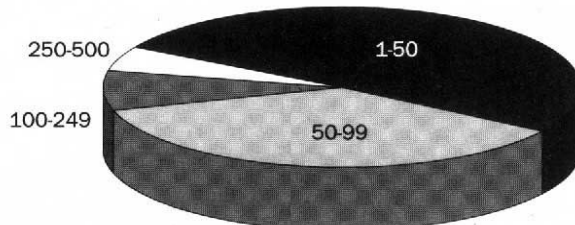
David Dickson

A profile of British biotechnology companies

Annual turnover



Number of employees



Source: Base Consultants, Ltd., 1992

route" for initial investors in British companies who, unlike their US counterparts, are required to stick with a company until it shows a profit.

Shortly before Christmas, the Stock Ex-

Research council takes stake in start-up

London. The Medical Research Council (MRC) has accepted a significant equity stake in a new company created to exploit research funded by the council in gene therapy technology. The company, known as Therexsys, hopes to develop a gene therapy technique that targets therapeutically effective genes to human cells without using more conventional vectors such as retroviruses and adenoviruses.

Both the MRC and the Cancer Research Campaign (CRC) — through its technology transfer arm, CRC Technology — have taken equity in the company in return for access to their respective research facilities, and in MRC's case for the exclusive rights to patents on research carried out at the council's National Institute for Medical Research

(NIMR) in London.

If the techniques prove successful, both organizations will be able to reap substantial rewards by selling their stakes either when a new round of financing takes place, planned for later this year, or when the company eventually goes public.

Therexsys was set up last summer with seed money from three British venture capital funds, Biotechnology Investments, Ltd. (the venture capital arm of the merchant bankers N M Rothschild), Schroder Ventures and 3i. It is being run from a science park in the northwest of England by Roger Craig, formerly professor of biochemistry at Middlesex Hospital Medical School in London, and until last year head of biotechnology for ICI Pharmaceuticals.

Therexsys hopes to develop an approach to gene therapy in which genes are accurately targeted and regulated. The technology could be applicable to both acute and, eventually, chronic therapy in a number of disease areas, including cancer, cardiovascular disease and the treatment of inflammation.

No details have been released of the precise techniques under development, partly because the company is still in the process of applying for some of the important patents. But they are thought to involve ways of inserting genes into cells using the ability of ligands such as monoclonal antibodies to target tumour and other diseased cells.

The MRC's contribution is based on research on the regulation of gene expression at NIMR by a team led by Frank Grosveld, head of the institute's laboratory of gene structure and expression. Therexsys will have an exclusive licence to the intellectual property, both existing and in future, arising from this work.

The practical application of the techniques will be explored by a research group headed by Mike Dexter, head of the department of experimental haematology at the Paterson Institute of Cancer Research in Manchester. Dexter is widely known for his work on the biology of the stem cells that give rise to blood cell lineages and for his interest in blood-borne tumours, which could be particularly susceptible to the ligand-mediated insertion of genes.

Craig has strong links to both the MRC and the CRC. While still at ICI, he belonged to a number of the research council's committees, including its neurosciences committee, its genetics policy group, and the directed programme committee of the Human Genome Mapping Project. He is also a member of the governing committee of the Paterson Institute.

MRC's past efforts to encourage technology transfer from its laboratories have been based primarily on negotiating licencing deals for work from its laboratories, and receiving income through the resulting royalty payments. That was the model for the biotechnology company Celltech, established with government help in 1980. But the company has yet to make a profit, and the return to the MRC so far has been disappointing.

Dai Rees, the secretary of the council, says that the MRC "has now realized the advantage of a mechanism which does not involve royalty deals, but in which we put in the knowledge, taking an equity stake in return, and other [investors] put in their capital".

The council already has a small equity stake in Cambridge Antibodies, based on work on antibody engineering at its Laboratory for Molecular Biology in Cambridge, and a similar stake in Somatogen, Inc, which works on artificial blood products. In contrast, an equity stake is a new idea for CRC Technology, which plans to return any profits to the CRC through covenants.

David Dickson

Congress asks universities to justify special funding

Boston. The chairman of the science committee in the US House of Representatives hopes to use US research universities as a wedge for reclaiming political power.

Representative George Brown (Democrat, California) has begun a study of the growing congressional practice of 'earmarking' federal dollars for large academic research facilities that have been neither requested by the relevant federal agency nor debated and judged worthy by the entire Congress. Last week Brown mailed a letter to 50 universities that received as much as \$10 million last year for such projects, asking them to explain themselves.

Brown has fought the growing practice (\$773 million was allocated to some 500 projects last year, triple the amount spent in 1990) on the grounds that it subverts the normal process of scientific review and because "it comes out of the hide of existing programs". But he admits that his campaign against so-called academic pork-barrel projects is also part of an effort to reestablish the importance of authorization committees such as his.

"The regular order of Congress requires universities to submit their projects to the appropriate agency and then to have that agency obtain an authorization to do what it wants", Brown explained last week during a press conference at the annual meeting of the American Association for the Advancement of Science (AAAS) in Boston. "Only

then should they turn to the appropriations process to get the money they need."

"Certain powerful members of Congress", Brown adds, have distorted that process by ignoring the authorizing committees and winning approval for pet projects at the final stage of the legislative process, when committees from each house meet to iron out the details of spending bills. The full Congress rarely has time to scrutinize the work of such conference committees.

For example, Senator Bennett Johnston (Democrat, Louisiana), is chairman of both the authorizing committee for the Department of Energy (DOE) and the appropriations subcommittee that funds it. The dual positions make it possible for Johnston to get money for certain projects without first having the money authorized, a luxury not afforded Brown and most other members of Congress. Perhaps not coincidentally, the energy and water appropriations bill is each year also loaded with such earmarks for universities.

Brown's questionnaire asks universities to describe the type of review their project has undergone and whether they sought approval from any federal agency before approaching Congress. University officials in the past have justified such projects by saying that there is no federal programme to repair and renovate academic research facilities and, thus, no alternative way to obtain such badly needed funds. **Jeffrey Mervis**

Massey says that NSF must tighten its belt

Washington. The National Science Foundation (NSF) must figure out a way to do its job despite insufficient resources, according to outgoing director Walter Massey.

Speaking last week with uncharacteristic frankness to a meeting of the National Science Board, NSF's governing body, Massey said that NSF staff must simplify the grant-making process and lighten its workload because "we aren't going to get the additional resources we need". That concession brings down the curtain on a promise made by presidents Ronald Reagan and George Bush to double the NSF budget within five years and could be a harbinger of the foundation's budget for next year, which will be released on 23 March.

Massey described for the science board the process that he began a year ago to develop a long-range strategy for the foundation. Although work on the strategy is still under way and may never be made into a public document, Massey said that he has asked senior NSF officials by May to draw up details of how to

implement it in five critical areas.

"We have to lower the barriers to interdisciplinary work", he said, "and we have to extend our resources by working more closely with other agencies. We need to better integrate education into each of the research directorates, and we need to do business differently to account for the fact that there is no realistic chance our [staff] budget will grow significantly."

The science board decided to spend a good deal of time in the next several months in discussing ways to implement the recommendations of the report last autumn by the Commission on the Future of NSF (*Nature* 360, 285; 1992). Although it would not be the first time that the science board has sought to play a larger role in defining US science and technology policy, board members believe that their chances of success have improved as a result of several recent reports on the subject and a greater receptiveness to such thinking by the new administration.

Jeffrey Mervis



Protesters in Munich march alongside a 'Trojan Oncomouse'.

Europeans protest against patent for Harvard mouse

Munich and Basel. The extent and power of the growing lobby against the European patent granted to the Harvard Oncomouse, licensed to the US company DuPont, was evident last week in the run up to last Saturday's deadline for formal oppositions. Co-ordinated demonstrations took place in ten countries, the Swiss government was petitioned to exempt Switzerland from patent coverage and, following an emergency debate, the European parliament voted overwhelmingly for an immediate revoking of the patent and for a moratorium on further patents for transgenic animals until a clear ruling has been established.

At least eight objections, called oppositions, to the mouse patent were filed by last Saturday's deadline, asking the European Patent Office (EPO) to revoke the Oncomouse patent. The patent, first filed in 1985, was granted nine months ago after a previous (successful) opposition was overturned.

Some 80 groups in Germany, representing at least two million people, banded together under the Munich-based umbrella organization *Kein Patent auf Leben* (No Patents on Life). A Swiss opposition was put forward by a similar group of 70 organizations. There is also opposition for the first time from one of Germany's 16 states, Hessen. British pressure groups filed an opposition

last month (see *Nature* **361**, 103; 1993).

The number of oppositions is significant given the high cost — DM1,200 (US\$800) — of filing one. The money goes towards operating the EPO, which unlike national patent offices is not supported by taxes. The oppositions will be considered by the EPO over the next year.

Last Thursday the European parliament, acting on a proposal from the Green faction, called for the EPO to withdraw the Oncomouse patent on the grounds that it violates the 'morality clause' within the EPO's convention and conflicts with parliament's amendment to the proposed directive on biotechnology under consideration by the Council of Ministers of the European Communities, which outlaws patents on transgenic animals designed to suffer. The parliament also voted 178:19 for a moratorium on further patenting. Although the vote has no legislative authority without the support of the Council of Ministers, it is expected to bring pressure on the EPO to change its regulations.

A petition with 7,300 signatures has asked the Swiss government to exempt itself from offering patent protection for genetically altered plants and animals. Switzerland is one of 11 European countries in which the Oncomouse patent is valid.

Alison Abbott & Oliver Klaffke

US Interior secretary considers a national biological survey

Washington. US Interior Secretary Bruce Babbitt is considering ways to consolidate biological research within his agency, using the US Geological Survey (USGS) as a model. At present, biological research activities are scattered across divisions that include the Fish and Wildlife Service, National Park Service, USGS and Bureau of Land Management.

Last week Babbitt met with staff from the existing divisions of the Department of Interior to learn what data on species and habitats already exist within his agency. A spokesman said that the Secretary "wants to get moving with [the plan] right now" in time to influence the department's budget for fiscal year 1994 to be presented in March.

Babbitt, a conservationist and former governor of Arizona, has talked about the need to focus federal research in plant and animal ecology and to strengthen the 20-year-old Endangered Species Act that comes up for reauthorization by Congress this year. In a television interview last week, Babbitt said that the conflict between loggers and environmentalists in the northwestern United States over the protected spotted owl is going to be the "test case" for the act.

A new biological survey could trigger a government-wide reorganization of environmental research, the goal of such groups as the Carnegie Commission on Science, Technology and Government (see *Nature* **360**, 699; 1992). The Clinton administration has not as yet addressed the issue, although last week it abolished the White House Council on Environmental Quality and replaced it with a new Office of Environmental Policy headed by 29-year-old Kathleen McGinty, a former Senate aide to Vice President Al Gore and a lawyer with an undergraduate degree in chemistry.

In the meantime, plans for a new US National Biodiversity Center at the Smithsonian Institution in Washington are on hold after George Bush failed in the final days of his presidency to sign an executive order approving the centre, as was expected (see *Nature* **361**, 197; 1993). Although the order had won the approval of several government agencies, it never emerged from the White House Office of Management and Budget and onto Bush's desk.

As it is now conceived, the National Biodiversity Center would be either a large central repository of data on plant and animal species or a network linking existing databases. An administration spokeswoman said that no decision has been made but that "we're obviously interested in the concept and are discussing it".

Tony Reichhardt

London meeting on British science

On Friday 19 March, there will be a meeting at the Royal Society, 6 Carlton House Terrace, London SW1, to discuss proposals for the forthcoming White Paper on the organization of British science. The speakers will be Sir Eric Ash (Rector, Imperial College London), Professor Michael Brady (University of Oxford), Dr Dai Rees (Secretary, Medical Research Council) and Sir Mark Richmond (Chairman, Science and Engineering Research Council); the meeting will start at 9.30 a.m. and end at 4.00 p.m.

Admission to the meeting will be by ticket, free of charge, obtainable from Mary Sheehan, *Nature*, 4 Little Essex Street, London WC2R 3LF. Coffee and tea will be provided.

There will also be a sandwich lunch for those who want it, **at a cost of £5**: please send a cheque made out to *Nature* with your ticket application. Tickets for both the meeting and lunch will be sent out during the second week of March.

Pay-as-you-go forensics could hurt British justice

London. The introduction of a "customer/contractor" relationship into the work of Britain's Forensic Science Service (FSS) may have made it harder for defendants to receive fair treatment, according to a report published last week by the Royal Commission on Criminal Justice.

The report, written by Paul Roberts and Chris Willmore of the law faculty at the University of Bristol, is based on a detailed analysis of 24 criminal trials in the Bristol region. It warns that the scope of the service has been limited by its need to satisfy a



"customer" — namely the police force — and that the need to do "what the customer wants" could affect its objectivity.

"Recent changes in the arrangements for payment for forensic science services could induce greater selectivity," the authors warn. "The danger is that forensic science evidence will reflect this selective approach to submission by the police, rather than giving a fair impression of the strength of the case against a defendant."

As an example, the report quotes a case in which the police asked a forensic scientist to compare a footprint found at the scene of a crime with shoes worn by one of several defendants in a court case. When the footprint did not fit, she contacted the police to ask if she had been supplied with all the relevant footwear and discovered that the footprints fitted shoes worn by another defendant, which had not been submitted for examination.

Roberts and Willmore describe this as a striking example of the way in which the choice of what items to submit can be selective. "It is most unlikely that the FSS would ever succumb to overt pressure, whatever the charging arrangements", they write. "Nevertheless, direct charging has perhaps increased FSS susceptibility to more subtle pressure."

In the past, the FSS was fully integrated into the Home Office. Since April 1991, however, it has been an independent, self-supporting body, and the police are now required to pay a fee for each item submitted for analysis.

According to one forensic scientist quoted anonymously in the report, the number of items submitted to her laboratory for analysis has fallen by 20 per cent since the introduction of direct charging. Even official statistics published by the FSS confirm a drop of 4 per cent in demand in its first year as an independent agency.

FSS officials say that the reduced demand may indicate that more cost-effective use is being made of the service. But another former member of the FSS, who was a defence witness in court cases, said that the new charging system is encouraging the police to submit only items thought to assist in securing a conviction. "Because they are paying on the basis of items sent to the laboratory, [the police] will be very selective in what they send," she is quoted as saying.

Criticism of the new procedures is backed by the Institute of Professionals, Managers and Specialists (IPMS), the trade union that represents 800 scientists working for the Home Office. "Direct charging can mean that the wrong decisions are being taken into account", Steve Jary of IPMS said last week. "Part of the problem is that the customer for the FSS should not just be the police or the prosecution services, but also the courts and the general public."

The report, which was prepared for the Royal Commission as part of its inquiry into the workings of the forensic science service following a number of recent miscarriages of justice, points out that many scientists working for the FSS feel that they should play a more active role in deciding how investigations should be carried out.

At the same time, the report emphasizes that all scientific evidence needs to be approached with caution. "Although science can have great utility in a forensic context, the risk of overreliance is obvious and attested to by the recent miscarriage cases," it says. "There can be genuine disagreement between forensic scientists just as there can be disagreement between nuclear physicists."

David Dickson

Correction

An official of the US Agency for International Development was improperly mentioned in a recent article on Indian protests against a seed gene bank under construction (see *Nature* 361, 291; 1993). The official, Joel Cohen, did not speak to the author of the article.

Strasbourg picked as permanent home for space college

Munich. Strasbourg has been chosen as the permanent home of the International Space University (ISU), which plans by 1995 to expand into an established graduate college.

The ISU, funded by government space agencies including the US National Aeronautics and Space Administration (NASA) and the European Space Agency (ESA), is temporarily based in Cambridge, Massachusetts. Founded in 1987 to promote international collaboration in space science, it began as a summer-school programme offering a ten-week course to some 130 graduate-level students, but the intention was always to become a permanent university.

The university will offer a single, diverse programme leading to a Master of Space Studies degree. The 10-month course will cover all aspects of space science, from physical sciences and satellite application to international policy and law, and will include a two-month practical project involving design of an international space mission. Initial enrolment is projected at 100, with double this number by the year 2000. Permanent and visiting faculty will number 20–40, and a PhD programme is a possibility. The ISU will be located within Strasbourg's Université Louis Pasteur.

Strasbourg was chosen over six competitors because of its international prospects and its strong academic reputation. The city and Alsace regional authorities have pledged the ECU28 million (US\$36 million) needed to get the project under way, and various sources, including ESA, some national space agencies and industry have pledged to meet the annual running cost of ECU6 million. City officials have six months to turn the pledges into reality.

Space scientists welcome the school's interdisciplinary and international approach to space research. At least 18 universities around the world have applied to become affiliate campuses, contributing to teaching on site and in their own establishments. Tony MacDonnell of the University of Kent's department of space studies believes that contact "with the internationally minded, young community of scientists that the ISU attracts" will be stimulating and that the students, with their broad education, will help to make space science more international.

Many of the students are likely to be seconded by industry. Those from academic backgrounds, says ISU spokesman Sterling North, are likely upon graduation to pursue management posts with space agencies.

Alison Abbott

Institute files for patents on first Japanese sequences

Tokyo. Japan has joined the rush to patent human genes on the basis of results of rapid cDNA sequencing. Last week, several Japanese newspapers revealed that Sagami Chemical Research Centre, a privately funded institute just outside Tokyo, has



Aerial view of the Sagami centre.

filed patent applications in Japan for about 60 human genes sequenced with a new cDNA technique.

The Sagami centre has applied for patents based on partial sequences of full-length cDNA clones for which some characteristics of gene function have been inferred. Its approach differs from that taken by the US National Institutes of Health (NIH) and Britain's Medical Research Council (MRC), which do not know the functions of many of the thousands of fragments of cDNA clones of human genes for which they are seeking patents.

The centre's move to patent human genes runs counter to a general policy among Japan's genome researchers not to patent the results of gene sequencing projects (see *Nature* 356, 181; 1992). But the Sagami centre is an unusual organization that depends on patent royalties to fund nearly half of its research (see below) and "the policy of our institute is that we should apply for a patent if we get patentable results", says Seishi Kato, the leader of the group at the Sagami centre.

Kato is one of the early proponents of the cDNA approach to human genome research, and he has developed his own technique to synthesize full-length cDNA rather than using cDNA fragments, the approach applied by NIH, MRC and many other groups around the world. But Kato has only a small semi-automated system of two US-made sequencers and two Seiko robots that at best can sequence 20 kilobases of cDNA a day.

"Sequencing is the rate-limiting step", says Kato, which explains in part why the number of genes for which the centre has made patent applications is so small compared with NIH and MRC.

Kato declined to reveal details of the patent applications. A reporter for the *Nikkei Shimbun*, Japan's leading financial daily, learned of the patent filings during a series of routine visits to Kato's laboratory.

Although unwilling to talk about the patents, Kato did say that the use of full-length cDNA rather than random fragments

allows him to target key parts of the clone for sequencing. He then compares these sequences with sequences on computer databases and is able to infer some of the characteristics of the proteins produced by the genes he has sequenced.

Kato says he is also able to produce proteins *in vitro* and *in vivo* directly from his full-length cDNA clones, something that in the case of random cDNA fragments can be done only after screening with probes. But his work on translation and characterization of proteins is still under way and the claims for gene function in the patent applications are based only on comparisons of his sequence data with computer databases. Even so, Kato believes that his partial sequences of full-length cDNA are patentable "even if gene function is unknown".

Kenichi Matsubara of Osaka University, leader of Japan's human genome project, does not expect researchers at Japan's universities and national laboratories to follow the Sagami centre's lead in patenting cDNA sequences. He says that there are no patent lawyers in Japan's government laboratories and this "tremendous defect in the present system" means that government researchers have to team up with private companies if they want to patent discoveries. But government regulations make that difficult, he says, adding that "unlike their US counterparts, not many Japanese companies seem willing to pay for patenting just sequences".

David Swinbanks

Research heads discuss impact of Framework

Munich. The heads of Europe's national research organizations have begun to meet informally to discuss the effect on their national efforts of a growing European science programme.

The European Communities (EC) will spend around ECU14.7 billion (US\$19 billion) in its 1994-1998 Framework programme. Most of this will be spent on applied research, but the magnitude guarantees that it will exert an influence on national programmes as well.

After meeting for the first time last month in Bonn, the heads of the national organizations want to continue talking about ways to interact with the EC and to share in the development of Europe-wide science and science policy. The next meeting is planned for next autumn in Britain.

In anticipation of the Maastricht treaty being signed, the research heads want to clarify how subsidiarity — a means to make decisions at the lowest possible level — will affect European and national research policies. They also want to rethink the role of the European Science Foundation.

Alison Abbott

Sagami keeps the mix boiling

Tokyo. The Sagami Chemical Research Centre, which has filed Japan's first applications for patents on human genes (see above), is a unique and, by Japanese standards, unusually dynamic research organization.

Established as a foundation in 1963 with backing from the Industrial Bank of Japan and 25 sponsors, including Hitachi, Nissan Motor Co., Tokyo Electric Power Co. and Nippon Steel, the centre has a rapid turnover of researchers. "Most only stay 5 to 10 years and the average age is only 32 or 33", says Akiyoshi Wada, one of the centre's directors.

Recruits are not offered permanent positions, an unusual policy for a country where lifetime employment is the norm. Rather, their performance is assessed after three years by the head of the institute and board of directors, and those judged good enough get a permanent post.

The centre has received several patents in the past two decades, most of them for organic chemicals, and 40 per cent of the centre's ¥1,400 million (US\$11.7 million) annual budget comes from patent royalties. Although the centre was established to "link chemistry and industry" and is dominated by organic chemists, its 100 researchers now include many biologists working on protein engineering, cDNA, enzymes and anti-cancer drugs.

D.S.

Catalonia increases spending but wants Madrid to help

Munich. Catalonia has lost its battle for autonomy from Spain in funding research, but it is determined not to lose the war over setting an agenda for science.

Last year, the Spanish high courts rejected Catalonia's argument that the constitution required the Madrid government to devolve responsibility for research to its autonomous regions. In response, the Catalan government has formed a high-profile research ministry and increased direct funding of local research by more than 10 per cent.

The struggle for complete independence in research has its roots in the Spanish constitution, drawn up in 1978. It gives the central government in Madrid jurisdiction over the "promotion and coordination of scientific and technical research", while leaving Spain's 17 autonomous regions responsible for setting "competencies on matters of promotion of culture [and] research". This legal ambiguity was clarified in the 1986 national plan for research, which chose to regulate research financing centrally.

Catalonia challenged the central government's refusal to transfer research resources directly to the regions. But last July the courts ruled in Madrid's favour.

The proportion of gross national product (GNP) allocated by Spain to research is one of the lowest in Europe — barely 0.9 per cent compared with an average of 2 per cent for the rich European nations. But industrially developed Catalonia argues that its re-

search potential has been stifled because it contributes nearly 20 per cent to the GNP but receives only 12–14 per cent of what the central government spends.

Fiercely independent, Catalonia dreams of becoming a major player in European research. It has promised to increase investment in research and its level of scientific

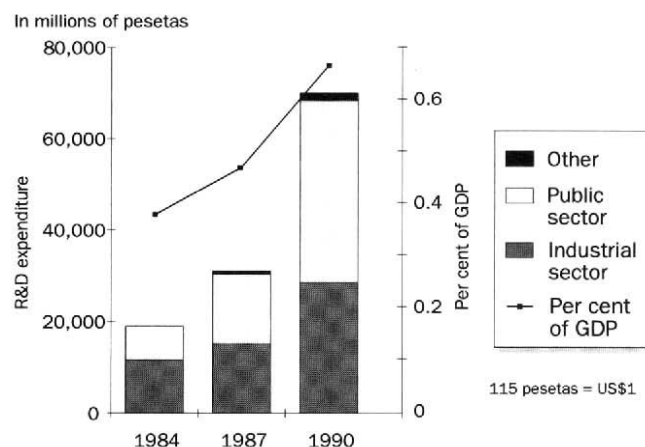
meantime, the only way to increase its research funding is to spend more of its own money. At the end of December, this is exactly what it decided to do. But the amounts fall far short of its own targets.

Responsibility for basic research has been transferred to a new department that combines responsibility for universities and university research. The department starts life with an increase of 13.5 per cent over last year, to Ptas3,521 million (US\$31 million). As well as funding investigator-initiated research, the department awards competitive grants within Catalonia's general plan for research and development. In addition, money for university research has increased by 7.3 per cent over 1992, to Ptas2,830 million, outpacing inflation, which last year stood at 5.4 per cent.

With the central government's contribution to research holding steady, the current growth is sustained with local money. Growth this year exceeds 10 per cent and may rise further. But the investment falls well short of parity with Europe's research elite.

Artur Bladé i Font, subdirector of research at the new ministry, says that the forthcoming general plan will dictate the speed at which the objective will be achieved; at the end of the first four-year plan, research spending is expected to reach 1.3 per cent of gross domestic product (GDP). "Who knows what will happen after that?" says Bladé i Font. "The government will have to decide if it can continue to make a 10 per cent increase every year." **Alison Abbott**

A growing commitment to research



personnel (it stands at 1.4 per thousand inhabitants, compared with an average of 2.2 for other European countries). Like most autonomous regions (except for the Basque country and for Nevada), it is obliged to give all of its income taxes to the central government, for redistribution according to established national needs, including research.

Catalonia is fighting to retain 15 per cent of the revenue from income taxes. In the

NASA seeks comments from Landsat users

Washington. Under orders from Congress, the US National Aeronautics and Space Administration (NASA) wants to hear what scientists think about its Landsat Earth satellite system. A questionnaire asks users of Landsat data about their experiences and solicits recommendations from the Earth science and remote sensing communities regarding resolution, frequency of coverage, data formats and other technical specifications for future Landsats. The deadline for requesting a questionnaire* is 1 March.

The gathering of public comments was required by last year's Land Remote Sensing Policy Act, which returned Landsat to US government control after eight unhappy years of operation by the quasi-private Earth Observation Satellite Company (EOSAT) (see *Nature* 359, 353; 1992). At the same time that a self-sustaining commercial market for US remote sensing data has failed to

develop, government and academic researchers have had to pay more for Landsat data than in the 1970s.

The new law, which gives NASA and the Department of Defense joint responsibility for Landsat, mandates that data users in the US government and affiliated institutions pay only "the cost of fulfilling user requests" for unenhanced data from Landsats 4, 5 and 6, with no markup in price by EOSAT. Landsats 4 and 5 — the two satellites currently in orbit — are both on their last legs. Landsat 6 is expected to be launched next summer.

The law directs that NASA and EOSAT should agree on the details of the new data policy — including pricing — by 30 September. One sticking point in these negotiations will be the effective date of the new, lower prices for government and affiliated users. The law says that it should be no later

than the launch of Landsat 7 — expected in 1998 — but also provides for an unspecified phased "transition period".

In the meantime, pressure on EOSAT to wean itself from its government subsidy has led to the company's first 'sale'. Current Landsat data prices range from \$200 for older, low-resolution images from the satellite's Multispectral Scanner to \$4,400 for a full-scene digital image produced by the Thematic Mapper (TM) instrument. Starting this week, the company will sell older processed TM scenes from its archives for \$1,500 to government and commercial users and for \$750 to academic researchers. The sale ends on 1 September.

Tony Reichhardt

* Questionnaires are available from Stanley R. Schneider, Landsat Advisory Process Coordinator, NASA Code SED, 300 E Street SW, Washington DC 20546. Fax: (202) 358-3098.

Drugged by tobacco

SIR — The tobacco industry continues to support research claiming that nicotine is not an addictive drug. Recently, however, 17 million Italian smokers suffered nicotine withdrawal syndrome as a consequence of an interruption of cigarette supplies during November and December 1992.

The production and distribution of tobacco in Italy was under direct government control until December when Parliament approved a bill converting the state factories into private companies. Fearing that they would lose certain benefits, the tobacco factory workers went on strike for more than 30 days, emptying the shelves of thousands of tobaccoists all over the country. Long queues of smokers formed outside tobaccoists when supplies arrived.

Once supplies were exhausted, people smoked whatever they could lay their hands on, including cigars and even cigarettes made from aromatic herbs and intended to help people to stop smoking. Some snatched lighted cigarettes from the lips of other smokers in the street. Many people crossed national borders or caught ferries to buy cigarettes in neighbouring countries. On international flights, duty-free cigarette supplies were soon exhausted and customs authorities enforced more severe inspections at borders and airports, a measure that aggravated the craving. Luggage from international flights was opened before customs inspection in the search for cigarettes.

As expected, there was an immediate boom in the black market. The price of a packet of tobacco rose tenfold. Social relationships were characterized by people talking about the want of cigarettes and smokers openly showed nervousness and irritability in all social environments.

This pattern of behaviour fits the definition of craving recently formulated by the United Nations International Drug Control Program-World Health Organisation (UNIDCP-WHO): "Drug craving is the desire for the previously experienced effects of a psychoactive substance. This desire can become compelling and can increase in the presence of both internal and external cues, particularly with substance availability. It is characterized by an increased likelihood of drug-seeking behaviour and, in humans, of drug-related thoughts."

The pattern of social and individual behaviour associated with nicotine withdrawal is therefore identical to the pattern of behaviour produced, for instance, during abstinence from heroin or any other drug of abuse. Indeed, the reinforcing properties of nicotine and of all other drugs abused by humans

have a common neurobiological substrate, independently of their primary mechanism of action. Thus the massive nicotine abstinence syndrome in Italy provides clear evidence that cigarette smoking is addictive, and deserves extensive study for its psychosocial and sociopolitical implications.

Zvani L. Rossetti

Glan Luigi Gessa

*B.B. Brodie Department of Neuroscience,
University of Cagliari,
via Porcell, 4,
09124 Cagliari, Italy*

Spanish science

SIR — Alison Abbott's analysis (*Nature* **360**, 502; 1992) of the situation of Spanish science is basically correct. However, as one of the postdoctoral fellows with the privilege of having a contract after three years of training in the United States, I want to make it clear that the length of these contracts is not really three years but is limited to the duration of the scientific project to which the fellow is assigned. Thus it is possible to have contracts of one year or less, unless the grant that sustains the funding for the project is renewed. This situation clearly adds more anxiety to the already complicated adjustments of returning to Spain.

As Abbott implies, lack of development of a biotechnological industry and the general absence of participation of public or private institutions (besides CSIC) in basic research contributes to the shrinking job market. Active participation of the public network of hospitals in basic research or restrictive measures to prevent the almost systematic promotion of insiders to permanent positions in Spanish universities would certainly boost the sagging spirits of Spanish scientists.

Santiago Lamas

*Universidad de Alcalá,
Campus Universitario,
Carretera Madrid-Barcelona, Km 33,600,
28871 Alcalá de Henares (Madrid),
Spain*

Luis Izquierdo

SIR — Your readers will be sorry to hear of the death on 10 December 1992 of the Chilean biologist Luis Izquierdo at the age of 64. Izquierdo was strongly opposed to the Pinochet regime in Chile (see L. Izquierdo *Nature* **253**, 5; 1975 & P. Cordero *et al. Nature* **308**, 310; 1984), for which reason he was dismissed in 1981 from the post he had held since

1964 at the University of Chile.

From January 1982 to the end of March 1983 he was supported by the Chilean Society of Biology, and, thanks to national and international pressure from the scientific community, he was reinstated as professor of biology at the University of Chile in October 1983.

D. G. Whittingham

*MRC Experimental Embryology
& Teratology Unit,
St George's Hospital Medical School,
Cranmer Terrace, London SW17 0RE, UK*

C. Barros

*Pontificia Universidad
Católica de Chile,
Santiago, Chile*

New DNB

SIR — Work on a *New Dictionary of National Biography* has now begun, funded by a grant from the British Academy. The *New DNB*, to be published by Oxford University Press, will expand the present *Dictionary of National Biography* (which began publication in 1885) from about 37,000 to about 50,000 entries, revising or rewriting the existing entries in the process.

An important part of this exercise is to discover the views of users about the present *DNB* and what improvements they think might be made in the *New DNB*, and to enable potential contributors to offer their services and to suggest new entries.

A questionnaire has been prepared that allows the respondent to give his/her views and to sign up. Anyone interested in the *DNB* and its future is urged to write to me at the address below to ask for one and then to fill it in.

Colin Matthew

(Editor)

*New Dictionary of National Biography,
Oxford University Press,
Walton Street, Oxford OX2 6DP, UK*

Primo Levi

SIR — I was commissioned in 1989 by Hutchinson to write a biography of Primo Levi, and I have now returned from six months in Italy conducting initial research. Although there is no authorized biography of Primo Levi, I do have cooperation from members of his family. If any of your readers knew Levi, I would be very grateful to hear from them. I am especially interested to see any correspondence, manuscripts or other documents relating to Levi's life, work and character. The greatest care will be taken of any such material.

Ian Thomson

*21 Gloucester Street,
London SW1V 2DB, UK*

Preserving diversity

SIR — It would be mistaken, as Rivera-Núñez and Obón-de-Castro point out (*Nature* 360, 291; 1992), to equate biodiversity solely with genetic diversity (measured in terms of genes, alleles, DNA), or to equate biodiversity conservation with the task of collecting and storing germplasm or DNA sequences. It is now established, as they point out, that 'real' biodiversity is the consequence of diverse interactions between individuals within biological populations, their environments and anthropic factors.

But Rivera-Núñez and Obón-de-Castro continue by criticizing the established, *ex situ* biodiversity conservation paradigm or, as they see it, the "armchair, molecular-biased approach". They favour wildlife preservation by monitoring of endangered species and the establishment of *in situ* reserves and they say that local landrace material cannot be stored without an understanding of the conscious and unconscious processes that led to their appearance and distribution. They go so far as to assert that conservation of cultivated plants and domesticated animals is dependent on the survival of the cultures in which they originated. This reasoning is ill-informed, fallacious and extremely dangerous.

The consequence of Rivera-Núñez and Obón-de-Castro's sophistry is that a false, antagonistic dichotomy is established between the two predominant approaches to biodiversity conservation, *ex situ* and *in situ* techniques. When either genetic or species erosion predominates and resources for conservation are so limited, how can fellow conservationists be so foolhardy as to promulgate an unnecessary, divisive argument that will undoubtedly detract from the crucial fact that it is our duty to conserve biodiversity in whatever ways are appropriate? Surely *ex situ* and *in situ* activities should be seen as and are in fact complementary techniques that are used by biologists to conserve native biodiversity?

It is a demonstrable fact that landraces can be conserved *ex situ* without knowledge of the culture that produced or distributed them. It is obviously desirable to obtain as high a quality of background information relating to the conserved material as possible, especially if the material is to be utilized as well as conserved, but it is not essential.

Rivera-Núñez and Obón-de-Castro point out that ethnobiological and cultural diversity is in serious danger of loss because of the encroachment of unadapted Western culture. This problem, and the associated genetic erosion, can-

not be underestimated. Ethnobiologists, anthropologists, sociologists and politicians can all work together to help aboriginal survival, combat the erosion of cultural diversity and aid conservation in the whole biosphere. This task is vast; conservationists will have a prominent part to play, but it is likely to be a much longer term goal. This goal, however, should not be formulated at the expense of shorter-term biological conservation goals, which aim to conserve crop-related germplasm *ex situ* for the benefit of all humanity. If we do not conserve germplasm now and wait for multidisciplinary agreement on how to conserve the entire biosphere, there may be little germplasm left to freeze and conserve. At this time, surely a broad, holistic approach involving *ex situ* and *in situ* techniques is called for, and is most appropriate, if we are to preserve biodiversity for future generations.

Brian Ford-Lloyd

Nigel Maxted

School of Biological Sciences,

University of Birmingham,

Edgbaston, Birmingham B15 2TT, UK

SIR — People continue to be surprised that there is no complete basic inventory of the variety of life on Earth. Some specialists champion the description of large numbers of new species on the grounds that this would facilitate the conservation of biodiversity through the recognition and comment on greater numbers of species¹⁻³. But the description of new species needs to be viewed within cost or political constraints. And there are reasons why the potential costs of describing new species have been grossly underestimated.

For the foreseeable future, description of all the world's species will remain impossible. So much can be demonstrated empirically by dividing the estimated numbers of species to be described by the numbers described per unit time: the result is a very large figure. The number of undescribed living insect species may be in the region of 9 million⁴, of which about 7,250 species are described each year, for example. But such calculations do not account for the fact that a proportion of the descriptions of 'new' species are subsequently recognized to pertain to previously described species. The proportion of past descriptions thus far synonymized commonly attains levels of 20 per cent or more among groups of insects. Because this problem is compounded through time, there are obvious serious financial implications⁵.

Plainly, although a complete global species inventory may be unrealistic, the

description of a sizeable proportion of those species is not. In the face of the extra effort needed to cope with the problems of synonymy, descriptive work will need to be carefully targeted. Already work in support of applied and theoretical studies in fields such as agriculture and medicine consume much of the available taxonomic resources.

We are concerned only with the resources left over from these projects plus any expansion that may result from recent reviews of the state of taxonomy. Taxonomic work for biodiversity conservation will continue to be severely constrained by the availability of appropriate expertise, and, without clear guidelines as to the conservationists' priorities, it will not be possible to provide an appropriate taxonomic framework.

Taxonomy, like other disciplines, is driven by the interests of individual scientists. Descriptive effort is thus widely dispersed, in every sense. By contrast, three approaches might target descriptive effort in support of the conservation of biodiversity: to concentrate efforts and resources upon a few selected groups, on a few selected areas or on a few selected systems. More communication between taxonomists and conservationists is necessary to meet these ends.

Kevin J. Gaston

Lawrence A. Mound

Department of Entomology,

Natural History Museum,

Cromwell Road, London SW7 5BD, UK

1. Soulé, M. E. *Ann. Missouri Bot. Gard.* **77**, 4-12 (1990).

2. Wilson, E. O. *Issues Sci. Technol.* **2**, 20-29 (1985).

3. Wilson, E. O. in *Biodiversity* (ed. Wilson, E. O.) 3-18 (National Academy Press, Washington, DC, 1988).

4. Hammond, P. M. in *Global Biodiversity, Status of the Earth's Living Resources* (ed. Groombridge, B.) 17-39 (Chapman and Hall, London, 1992).

5. Gaston, K. J. & Mound, L. A. *Proc. R. Soc.* (in the press).

Due credit

SIR — Perhaps the present Vatican deserves more credit for rehabilitating Galileo than you give it (*Nature* 360, 2; 1992). From my own observation, I believe the dogma of the Church is much influenced by who the Pope and his advisers happen to be at any given time. In seeking to understand the natural explanation for the evolution of life, from repeatable and repeated experiments, the present Pope is ahead of many in the scientific community. With rehabilitation of Galileo and with openness to modern views of Genesis, the Vatican under Pope John Paul II appears to signal a more informed relationship between science and religion.

Sidney W. Fox

Mineralization Center,

University of South Alabama,

Mobile, Alabama 36688, USA

Telephone book of life

SIR — There are many different kinds of information systems such as libraries, computers, telephone books and the human genome. Would it help to understand the organization of the genome into coding and noncoding or 'junk' DNAs to the similarity of 'junk' DNA to information that we encounter every day in other information systems?

Familiar information systems have common characteristics. One common feature is that each information file occupies a unique location in three-dimensional space, and access to a file is obtained by using codes. This means there is considerable access information contained in the architecture of the system itself. Think how difficult it would be to find a book in a library if all the books were piled in a large heap (even if they all had the proper catalogue numbers printed on them) or to find a telephone number if all the numbers were printed alphabetically on a single paper tape. Therefore, for ease of use, information systems usually consist mainly of the hardware architecture and access codes required to find information quickly and efficiently.

If we think of the cell nucleus as an information system consisting mainly of hardware (chromatin and nuclear architecture) and access codes (transactivator binding sites), then several characteristics of the nucleus and genome might become a bit more understandable because they may follow the same rules as other information systems.

A telephone book consists of hardware architecture in the arrangement of characters on its pages and of access codes for names and addresses. Because the only 'coding' information in the book is the telephone numbers themselves, any telephone book is composed mainly of noncoding sequences required to find those numbers — you don't use a telephone book to find names; you use the names to access telephone numbers. Promiscuous access motifs are found throughout any phone book — such as John, Mary, Dr, St or Ave — and they can mean different things depending on their location — as for Dr John Johnson of 63 Upjohn Dr.

To find hundreds of mindless repetitions of the same five-letter motif, one need only to go to Smith, A. and read down the left margin to Smith, Z. The access codes even work with multiple mismatches to a consensus spelling — such as Chas or Jos. User-specific genes and pseudogenes are common in a telephone book — you get the numbers of all the butchers in town in your book even if you're a vegetarian and you get numbers for people who have died or

moved away. Certainly the cell nucleus and the human genome it contains comprise 'the book of life'; not a novel to be read but rather a telephone book to provide correct information to individual users when they need it.

If anyone is uncomfortable with the idea of huge amounts of noncoding and junk DNA within the information storage medium of the human genome, I suggest four thought experiments which should indicate that noncoding and junk information are integral parts of any information system:

(1) Compile a complementary DNA library of a telephone book — to include only the telephone numbers (without the noncoding names and addresses next to them) that will be needed during the next year.

(2) Compile a personal computer floppy disk with all the information necessary for word processing, but where the 'noncoding' system and operating programs require less memory space than the 'coding' manuscript files.

(3) Scan the hard disk on your own personal computer for 'junk' — old or outdated manuscripts and data files that will probably never be used but that fill up the disk 'just in case they may be useful some day'.

(4) Devise a logical method to scan the shelves in a library and remove all the books that will never be used in the lifetime of that library.

I believe that we can begin to obtain insights into the structure and function of the cell nucleus and human genome by assuming that it operates by rules similar to other information systems we use daily. If we look, we can see noncoding and junk information all around us.

John W. Bodnar
Chemistry Department,
US Naval Academy,
Annapolis,
Maryland 21402, USA

European funds

SIR — My impression of the European Commission (EC)'s Human Capital and Mobility programme is different from that of Lennart Philipson (*Nature* 360, 102; 1992). My experience is that of coordinating one of the Joint European Projects within the EC TEMPUS programme; although focused primarily on university teaching, it also has a large component of trans-European academic mobility.

The EC's recent involvement in funding of academic mobility has indeed completely changed the funding scene for European researchers. While fine-

tuning of the procedures may be required, the Human Capital and Mobility programme is the best news European science has had for some time.

EC funding is available on a competitive basis to almost any centre within the European Communities. This is of particular interest to interdisciplinary scientists whose programmes might be judged to fall outside the scope of specialist organizations.

Moreover, in the past, more often than not, an award of a travelling fellowship has spelled financial disaster to a successful postdoctoral fellow. Better personal support will in itself assist the mobility of high-calibre researchers.

Inevitably, this new source of funds dwarfs in scale the present funding organizations. But nothing prevents individuals who are members of smaller organizations from applying for funding from the large EC programmes. Indeed, most people are familiar with a polite letter from the granting body saying "we hope you can obtain funding from another source".

Now, a realistic alternative source of funding exists. Notwithstanding procedural adjustments that still need to be made, it is certainly excellent news.

Marek H. Dominiczak
Department of Pathological Biochemistry,
University of Glasgow,
Western Infirmary,
Glasgow G11 6NT, UK

Centre move

SIR — I believe that now is the appropriate time to reconsider and correct the term "Eastern Europe", often used in *Nature*.

This term was developed after 1945 to represent political rather than geographical realities in Europe and was used for countries under Soviet domination, despite their being in Europe. A few years ago the political situation changed drastically and many countries were liberated from the presently extinct Soviet Union.

The old term "Eastern Europe" now represents a nonexistent political situation in Europe and should be changed to a more precise geographical term. East Germany is no longer in Eastern Europe, while Poland, Czechoslovakia, Austria, Hungary and other countries of this region are Central European countries: Belarus, Ukraine and part of Russia are East European countries.

J. Barankiewicz
Gensia Pharmaceuticals, Inc.,
11025 Roselle Street,
San Diego,
California 92121-1204, USA

□ The point is well taken. — Editor, *Nature*.

How will Britain run its science now?

Britain is a glutton for the reorganization of its research enterprise. There have been two major upheavals in the past 30 years, in the early 1960s and in 1972. Soon there will be a third.

THE appointment last April of Mr William Waldegrave as Chancellor of the Duchy of Lancaster in charge of the newly established Office of Science and Technology (OST) within the Cabinet Office was widely welcomed. Waldegrave is an energetic and intelligent person. He had been a protégé of the late Lord (Victor) Rothschild in the early 1970s, when Rothschild was the head of the Cabinet Office's think-tank; he is knowledgeable about and even sympathetic to science.

• Waldegrave has made the right noises from the beginning, offering dialogue with the scientific community and a willingness to listen. This promise he has kept, receiving delegations and even listening to what they have said, visiting laboratories and talking to the people there. Unlike his predecessors, Waldegrave seems to welcome meetings with academics, both privately at Oxford and Cambridge colleges and in cavernous lecture halls. He has all the makings of a good guy.

That is one reason why his well-wishers hope that his first major project, a White Paper (policy document) on the organization of British science, does not blow up in his face. The project was announced quite early on, and was followed by a remarkable exercise in consultation. About 800 organizations and individuals offered opinions about the contents of the white paper. Some of these deal with narrow issues, others plead special causes (such as the importance of physics or taxonomy) and still others amount to virtually comprehensive prescriptions for the organization of public support for science in Britain.

How, with all that advice, could a minister go wrong? There are several dangers, of which perhaps the chief is that most of the advice that Waldegrave has been offered is surprisingly radical. Even the Royal Society advocates the creation of a new research council operating exclusively in support of basic research, and in the physical sciences in particular. The government's Advisory Council on Science and Technology (ACOST) goes much further, advocating a kind of separation of powers between the recipients of government research support (called "research providers") and the organizations that supply the funds, research councils for example. The temp-

tation to produce a white paper that is more radical than the circumstances warrant will be great.

A more serious danger is that the theme of research for wealth creation, which Waldegrave intends to be a principal theme of the white paper, has prompted much influential but simplistic advice. Waldegrave's white paper will either have to accommodate these pleas for close linking between all publicly supported research and wealth creation, or it will have to give persuasive reasons why a substantial effort in undirected research should continue.

A further difficulty is that the welter of advice says very little about what some

pean Communities's growing research programmes. That is as it should be. No doubt the white paper will deal with this point. But will it risk advocating more imaginative ventures, cross-frontier grant-making or the pooling of publicly supported research on essentially identical research goals, from weights and measures to the improvement of varieties of grass for herbage? We shall see.

What follows is an opinionated commentary on some of the issues certain to be dealt with in the white paper. It is intended primarily as a background paper for two public meetings in which *Nature* has a hand, one (in collaboration with the Edinburgh International Science Festival) at Edinburgh on 19 February; a report of that occasion will be available for the second meeting at the Royal Society in London on 19 March. *Nature's* own opinions appear in italic type (*like this*) for the convenience of those who know what they are (see "Manifesto for British science" *Nature* 353, 105-112; 1991) or who disagree with them. Although what follows is concerned with domestic British matters, it is hoped that it will be of some interest, perhaps macabre, to those working elsewhere.

Rex Features

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Waldegrave: much advice to mull over.

(including *Nature*) regard as one of the most serious problems now facing British science — the impoverishment of salaries in British basic science relative to those in other professions, and the fragmentation of traditional career structures. These circumstances encourage the loss of skilled people from research (by emigration and otherwise) and discourage the recruitment of able people into science. Waldegrave's Office of Science and Technology, through its dependent research agencies and the publicly supported universities, does not need legislation to influence research pay, but a few paragraphs on the subject in a white paper would encourage the troops and help keep the Treasury at bay in the future.

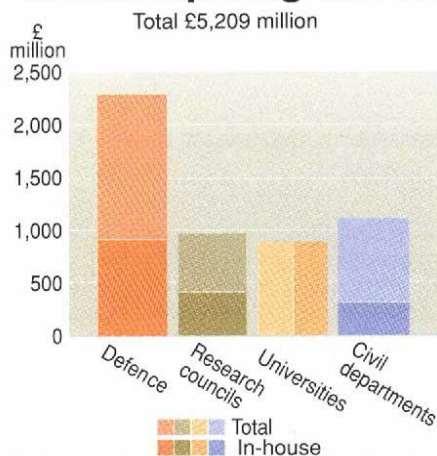
The advice to Waldegrave also skimps on Britain's European connections and their consequences for research. The Royal Society is perhaps the most explicit, urging (among other things) that Waldegrave's office should take a firm grip on Britain's representation in Brussels on the management of the Euro-

Organization

To traditionalists, Waldegrave's appointment last year was a fond echo of the brief spell during the Macmillan and Home governments of 1960-64 when Lords Hailsham and Jellicoe, members of the British cabinet with the formal title of Lord President of the (Privy) Council, were given responsibility for science. More directly, the appointment was a belated response to the argument of the House of Lords Select Committee on Science and Technology in 1981 (and on several succeeding occasions) that there should be a minister with responsibility for government-supported research. It was essential to that proposal that the minister should not be a full-time Minister of Science, not least for fear that otherwise he would be mistaken by his colleagues for a narrow lobbyist, even a captive of his constituents. (Waldegrave's other responsibilities include the Prime Minister's Citizens' Charter project and the Civil Service.)

The administrative changes accompanying Waldegrave's appointment included the transfer to OST of what used to be the Science Branch of the Department of Education and Science (still the

Total R&D spending 1990–91



Defence-related research is dominant.

channel for university funds), which thus also inherited the Advisory Board for the Research Councils (ABRC). That endows OST with oversight of the activities of the research councils, of which there are at present five.

One crucial question for the white paper is that of what other authority OST should have. Many of the organizations offering advice suggest that OST should become the branch of government concerned with setting research priorities across government. ACOST says it "should develop a coherent national strategy", the Royal Academy of Engineering that "the possibility of putting responsibility for other government research funding under OST" should be explored and the Royal Society that OST "should develop, continuously, a strategic vision" of what science and technology can accomplish.

Nobody will dispute that the British government needs a means of assessing the relative value of the components of the £5,000 million a year it spends on research and development, less than a fifth of that through the research councils, more than twice as much through the Ministry of Defence and the remainder through ministries such as the Department of Trade and Industry (DTI) or the Department of the Environment. As things stand, there is no effective means of carrying out that task. Indeed, there never has been.

It is true that the Rothschild reorganization of British science (outlined in November 1971 and ordained, with crucial dilution of purpose, in July 1972) included as a crucial element a system in which all research-based ministries would appoint chief scientists who would have looked to the government's chief scientific adviser as their *de facto* controller. Rothschild had written a job description for a post he would willingly have taken himself, but which he was not invited to fill. In the event, the departments would not tolerate such interference.

We shall be able to judge from the white paper how much of this authority Waldegrave has won. In the event, DTI will have been his chief adversary, and on several grounds: the department has hitherto looked after British representation at Brussels on research (which the Royal Society would transfer to OST). It is also the public sponsor of manufacturing industry, whose well-being many respondents wish to enhance by publicly sponsored applied research and which is at present headed by a powerful minister (Mr Michael Heseltine) whose civil servants will urge him to fight for their own independence.

Nature's view is that, because of the nature of government, efficient coordination of all publicly supported research is an impossible task, that in a sophisticated industrial economy the definition of a "research strategy" is likely to take as long as the length of time for which any strategy will be persuasive, but that a central office, adequately staffed with professional people and well (if informally) advised, could do much to nudge government departments and groups of industries in useful directions. That is how Japan's Ministry of International Trade and Industry works. M. Hubert Curien, minister of research in Paris for the past decade (excluding the brief tenure of the Chirac government) has single-handedly sought to do much the same for France. If Britain aspires to a government function of this kind, executive responsibility should rest with DTI.

Those who would wish on the stripling OST the task of devising a national strategy in research have, in any case, overlooked the basic housekeeping tasks that OST was saddled with by last year's reorganization. The Science Branch of the Department of Education and Science was responsible for £1,000 million of spending each year by the research councils. However this budget is reorganized, funds of that magnitude will still have to be administered, which is not a negligible task.

Indeed it is inevitable that, under the new arrangements, OST will be more directly involved than was the Science Branch in the scrutiny of its dependants' annual spending plans, while the Royal Society has cogently argued that OST should itself have funds of the order of £100 million a year to support "activities transcending the responsibilities of individual departments" and for providing matching funds for projects partly supported by the European Communities. There will be a lot to do.

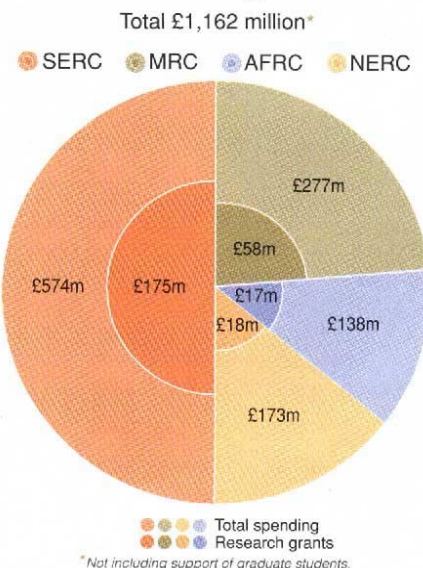
The central responsibility that could realistically be discharged by OST falls short of the "coordination" and "planning (of research)" that many respondents have urged on Waldegrave. Since the Rothschild reorganization (and, in prac-

tice, since the change of government in 1970), Britain has lacked an effective means of identifying and dealing with major questions of science policy. Before that, between 1964 and 1970, a body called the Council for Science Policy advised the Wilson government on matters such as the emigration of British scientists and the place of high-energy physics in the pattern of civil science spending (and was free to publish its advice). Before that, an Advisory Council on Scientific Policy did the same job more thoroughly if less adventurously.

Despite Rothschild's plea that government departments should be equipped with chief scientists well supported enough to be intelligent and perceptive customers for research, it has since been left to civil servants to decide what should be done to make better civil use of defence research efforts, whether Europe's farm surpluses meant that the volume of agricultural research should be reduced (the decision could have gone the other way) and how best to provide public research support for biotechnology or information technology.

The doctrine that government should not support "near-market" research was similarly the creation of a civil servant (Dr, now Sir, John Fairclough), the second of Mrs Margaret Thatcher's science advisers. Should not industry, instead, commission and pay for research intended to yield industrial benefits? The doctrine was used in the late 1980s to justify a retreat from applied research in government laboratories and a steady reduction of DTI's research spending. Much of the advice to Waldegrave asks that the doctrine be repealed. That seems inevitable, although repeal would be easier if the "near-market" criteria had ever been publicly defined.

The science budget 1993–94



'Core spending' for research councils is on academic research grants.

Whatever may be OST's housekeeping function, it is the only agency of government placed centrally enough to ensure that such questions are quickly identified and thoroughly discussed. There is a host of issues that need attention, ranging from the evaluation of the schemes for research support invented by the research councils in recent years and the adequacy of the arrangements for supporting academic research projects by overhead payments (new this year) to grand questions such as the scale of continuing support for defence research, the conditions of employment in British science and the opportunities for more effective collaboration on a European scale.

OST could hope, in time, to be regarded as a leader of the British research enterprise, perhaps in the sense intended by the Royal Society's phrase "strategic vision", but only if it has the right to examine (not necessarily to decide) questions touching the whole of government spending on research — and if, meanwhile, it can tackle such of the research community's discontents as fall squarely within its terms of reference.

Advice

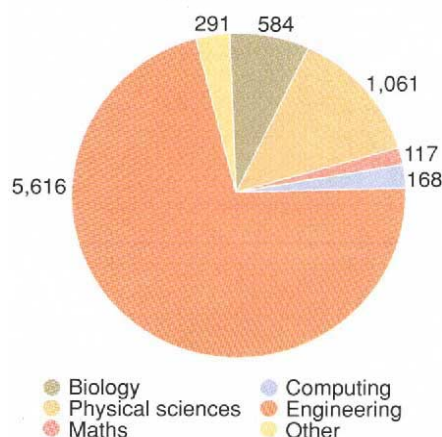
Waldegrave's invitation of advice asked for comments on the government's advisory apparatus in science. He has been given plenty, much of it predictable.

Opinions are divided on whether the Advisory Board for the Research Councils (ABRC) should continue; ABRC itself foresees the continuation of a board to recommend the allocation of funds between the research councils (its own chief function), but others hold that the job could be done by OST now that it exists. (The Science and Engineering Research Council, which would prefer abolition, makes the damaging point that the distinction between advisory and executive functions has been blurred in recent practice.)

In whatever form the research councils themselves survive the white paper, the assessment of their annual claims on public funds is hardly a task the officials concerned would relish; research council heads, after all, are powerful fellows. Suggestions that the scope of the individual research councils should be modified to reflect more precisely defined "missions" (for which there are several recipes) can only make the comparison of their claims more difficult — not so much the comparison of different kinds of cheese but of different kinds of chalk as well. Moreover, if the research councils are no longer *ex officio* members of an assessment board, the degree of self-restraint at present stemming from their knowledge that fellow-members will mock over-ambitious claims will presumably be removed.

A prudent OST will therefore need

Recruitment into R&D of first graduates 1990



Engineering, added as the 'E' in SERC, is touted as meriting its own 'ERC'.

expert advice on its housekeeping functions. How might that be done? Although the research councils will be vitally interested in the outcome, there are other interested parties — higher education, industry, even other government departments. On the assumption that research councils have clear terms of reference within which their existing programmes of research support have been justified, the annual adjudication of their claims will mostly consist of an assessment once a year of their proposed innovations.

In the circumstances, there is a strong case for breaking with the traditional concept of a standing committee meeting once a month to carry out an annual task. An ad hoc assessment panel putting in perhaps a month's hard work would seem to meet the need more adequately. It would be beneficial in such an arrangement that there should be a substantial changeover in membership from one year to the next. This would also be an opportunity for Britain to venture into a field from which it has too long shrunk — the engagement of technical people and industrialists from overseas to bring international yardsticks to the assessment of British research and development.

For the rest, there seems general agreement among Waldegrave's respondents that a committee such as ACOST should continue, with oversight of all government research and development. (ACOST would call it NACOST, where "N" stands for "National".) Such a body would indeed be necessary if OST were to have the authority to enquire into matters affecting other government departments.

This near-unanimity would be the more compelling if the existing committee had been as great a success as ACOST's advice to Waldegrave suggests. But ACOST has not been outstandingly more effective than its predecessor

ACARD (the Advisory Council on Applied Research and Development). For one thing, its work-rate has not been particularly commendable, no doubt because many of its members are captains of industry, or are otherwise busy with the administration of science. The topics it has chosen for its occasional reports, while interesting, have rarely been central issues. Publicly, at least, it has not dealt with the "near-market" criterion, the transfer of the nuclear energy and defence research establishments to nominally independent agencies or the question (prominent in Waldegrave's invitation of advice) of why Britain seems less successful than its competitors in the exploitation of research.

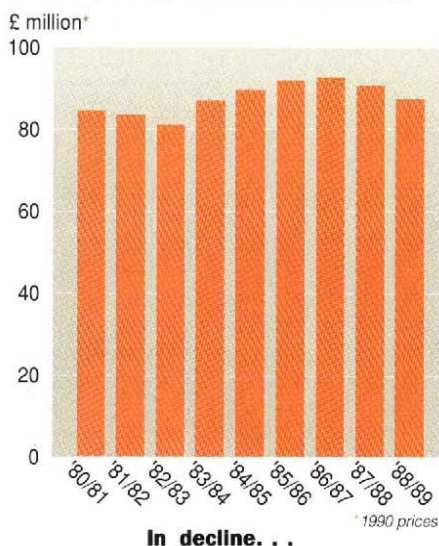
The Royal Academy of Engineering, in its advice to Waldegrave, would perpetuate those weaknesses. The academy recommends a committee of some "8 or 10 members including company chairmen or chief executives from industry and senior academics" with "six or seven short meetings a year to provide advice on policy issues and priorities".

There is also some confusion about the sense in which this overall committee is and should be independent. ACOST sets great store by the fact that its chairman is neither a minister nor a civil servant, and by its meetings twice a year with the Prime Minister. But none of that guarantees that a committee is free to take up questions it considers to be important, let alone to advocate proposals that the government and its officials have not thought of first; for that, it needs an independent secretariat as well.

The advice to Waldegrave also overlooks one of the most important constructive functions this grand committee could perform. Part of the frustration of the scientific community in Britain now derives from the sense that even serious discontents with the management of research were never considered seriously within government. The government would be better served if the members of its overall committee were not merely distinguished but accessible to the community, and able to act as lightning conductors for dissident opinion.

One model is the US President's Science Advisory Committee until its reconstitution by President Richard M. Nixon in 1971: that succeeded in combining the giving of advice on confidential matters (such as nuclear weapons development) with that of serving as a channel of communication between the White House and the scientific community. Nearer home, the Advisory Council on Scientific Policy of the 1950s served just that function in Britain. Necessary conditions for success in this regard are that the advisory committee should set its own agenda and that its members should see themselves, in part, as representatives of the scientific

Research council expenditure on postgraduate training



community — and should have the time to discharge that responsibility.

Providers

The most radical advice to Waldegrave is that of ACOST, which urges that “the customer-contractor principle needs to be implemented more effectively”. The reference is to Rothschild, who argued that government departments’ chief scientists should regard themselves as customers for research, which they would then commission from research councils (the contractors) and possibly other sources.

ACOST, indeed, would split the present research council system in two orthogonal directions. Potential contractors would be separated from the providers of funds, which would be prevented from “owning” research institutes. ACOST cites as one of its anxieties about the present system the inhibition of potential competition between research providers (institutes and research groups) because of their special relationship with, say a research council. It would constitute all public research laboratories, whether managed by research councils or government departments, as “free-standing” establishments.

The orthogonal direction is the proposed separation of “curiosity-driven” and “mission-oriented” research: ACOST recommends two new research councils, called CASK (Council for the Acquisition of Scientific Knowledge) and CMOR (the Council for Mission-Oriented Research), which would operate through a number of programme directorates in fields as different as health, electronics and production engineering. In ACOST’s opinion, government departments would continue as at present to make arrangements with industrial companies and research providers to carry out collaborative program-

mes of research.

These proposals, thus starkly described, will unsettle all those used to the pattern of research councils that has grown up over the past 30 years. It is therefore important that there are strong echoes of ACOST’s view in other influential advice to Waldegrave.

Thus the Science and Engineering Research Council (SERC) asks that the research councils should be free to make research grants to all applicants — that there should be an “open market” in research — and that there should be a “clear distinction” between purchaser and provider. It recommends four research agencies dealing with environment and land use, medicine, the needs of industry and a “cultural” research organization; the last is taken as synonymous with the maintenance of the “intellectual knowledge base”.

ABRC, in similar vein, advises of the need for “market testing” in research; that is a current code-word for the process of finding out whether some other organization can do a given job more cheaply. ABRC would have five research research councils covering the natural sciences, but would rationalize support of research in social science and the humanities. The Royal Academy of Engineering, on the other hand, would have an Engineering Research Council.

Significantly, none of the organizations advocating restructuring of the purchasers’ terms of reference mentions, even for illustrative purposes, what it considers should be the budgets of the new organizations under the new regime.

Readers of these opinions may also think it odd that so little of the advice to Waldegrave, even that which insists that research policies must be determined by the inclinations and enthusiasms of researchers (presumably intended by the frequent use of the phrase “bottom-up”), considers how, and on what conditions, researchers and their students will willingly participate in new arrangements. Their assent is taken for granted. That is both unwise and tactless.

Nature’s opinion is that the social function of “curiosity-driven” or academic research is not primarily the “advancement of scientific knowledge”, enlivening though that may be, but the education of able men and women in the knack of solving previously unsolved and even un-confronted problems; that the fields in which these usually young people are educated in research are not those in which they will eventually contribute to the gross national product; and that their number, which has begun to shrink, may catastrophically collapse if they continue to be offered during their formative years stipends roughly a third of the average national wage.

Which of the organizations in the new

scheme of things should be responsible for this crucial function? Most of the advice to Waldegrave would distribute the responsibility (as now) among several research councils. There is a strong case for a different solution.

Neither the standard British three-year undergraduate course nor the four-year Scottish course is a sufficient preparation for a career as an independent and creative scientist. Rather, practical experience of solving previously unsolved problems is now an essential part of the education of working scientists, both in industry and in academic life. And Britain needs more of these people if its economy is to prosper. The advice to Waldegrave in which reorganization of the research enterprise is offered as a means of enhancing wealth creation skips too lightly over Britain’s urgent need of more technical skill.

Within broad limits, the particular problem on which a graduate student is engaged may not be crucial; a person working on a problem even in astrophysics will, if he or she succeeds, know what blend of artifice and persistence is required to learn a little more about the natural world, and will at the same time become practised at techniques that can be applied elsewhere.

The apparently academic study of the conformation of the surfaces of solids is in many ways a better preparation for a career in the semiconductor industry than an apprenticeship at a manufacturing plant; the breadth of understanding is a guarantor of future flexibility, while the research costs will be less. In fields such as molecular biology, where the gap between basic and applied research has, for the time being, vanished, the research projects on which graduate students are engaged are almost literally irrelevant to their future usefulness in industry.

These considerations imply that basic research should continue to be the chief substrate of young people’s education in research, and that academic institutions and basic research laboratories should be its chief providers. The novelty is that the assumption that a PhD course is usually a preparation for an academic career must be abandoned. Education in research is British industry’s chief lack (outside the successful pharmaceutical industries).

Nature would therefore like to see a Natural Sciences Research Council whose chief responsibility is the recruitment of young people into research, and which would then support their education in research by means of research grants to established researchers, often (but not exclusively) at universities (including the ex-polytechnics). A system for the pre-qualification of graduate students (as in France) should be introduced; better working conditions would be necessary to tempt young people over this extra hurdle.

New ways with the solar nebula

Was light from the primaeval Sun the agent of the differentiation of the planets as we know them? The density was probably too great for ordinary radiation pressure to suffice, but there are ways around that difficulty.

WHETHER or not the general opinion of the origin of the Solar System is more or less correct, the simple view that everything began in a contracting nebula of roughly homogeneous chemical composition fails to carry quite the conviction of a well polished model. There are too many loose ends, such as the time-course of the evolution of the nebula in the early stages, before the Solar System came to resemble what we see today. With time, the loose ends will no doubt be tied together.

But there are also anomalies, departures from the expectations suggested by the simple model, which are more worrying, for just the reasons that a few grains of sand in an otherwise efficient lubricant may destroy the engine of a motorcar. The high mean density of Mercury has always been a bother. The circumstance that the Earth, Mars and Venus, similarly placed and roughly of the same mass, should have such very different atmospheres demands an explanation requiring a special assumption for each planet (such as life on the Earth, for example).

Part of the trouble is that knowledge of the Solar System has been accumulating too quickly for theorists' comfort. A few decades ago a then colleague, an ingenious, voluble and eccentric Ulsterman called W. H. ("Bill", as in Clinton) Ramsey, was able to fit on a smooth curve the known characteristics of the planets — mass, density and hardly anything else — by starting with the assumption that each had begun as a blob of solar nebula which had then been differentiated under the influence of the Sun. By the time he died (too young) in the late 1950s, Ramsey was already being overwhelmed by inconvenient facts.

Those who would make sense of the variations between the properties of the planets in the Solar System by starting with an undifferentiated solar nebula have a much more difficult task than Ramsey's: there is so much more to account for. The history of the Sun can be reasonably well inferred from what the stellar modellers have done, of course.

The Sun would have begun to form from the nebula by its collapse under its own weight, and its central regions would have been progressively heated by the energy thus liberated. So the outer regions of the collapsing nebula would have been exposed to an increasing outward flux of radiation of increasingly high frequency, infrared radiation first and radiation not very different from what we now see when nuclear reactions were switched on in the Sun's core.

Even so, it is far from clear whether the

material from which the planets eventually formed was prevented from becoming part of the Sun by its inertia (which would be a function of the speed of collapse) or whether it was actively kept out by, say, radiation pressure (to use that term loosely). For stellar modellers, that is a question for a second or third approximation in understanding the process of star formation, but for those who would understand the history of the planets, it is a central question. Would there have been enough of the nebula left over to form them, for example, especially when the formation of the inner planets, rich in heavy elements as they are, would have been exceedingly wasteful of nebular material?

It is also clear that the process of differentiation and segregation in the remnant of the nebula must have continued long after the Sun became a star. To a first approximation, the preferential loss of light elements from the inner Solar System may be accounted for simply by the temperature on the surfaces of the inner planets (which is what Ramsey used to do).

But there are snags, one of which is the variation of the ratio of deuterium to hydrogen from the inner to the outer planets. These data, which have become accurate only because of planetary spacecraft, are remarkable. On Venus, the ratio (by atoms) of D to H is 1.6 per cent, roughly 100 times greater than on the Earth, but in the outer planets it is not very different from the cosmic value, which is a further order of magnitude less.

Two Russians, S. N. Atutov and A. M. Shalagin, appear to have provided a way out of this conundrum. Their conclusion is that the variations of the D/H ratio cannot be explained by the difference of the velocities of hydrogen and deuterium atoms in the upper atmospheres of the inner planets, even though that effect is in the right direction. (The Maxwell distribution of velocities at a given temperature scales inversely with the square root of the mass, so that hydrogen is lost more easily than deuterium).

Nor, sadly, will radiation pressure in the ordinary sense suffice. Although atoms are impelled away from a source of radiation by the momentum of photons they may absorb, so that atoms that preferentially absorb radiation from a source are preferentially driven away from it, that mechanism is effective only when the density is small enough for collisions to be rare. By definition, that cannot have been the case in the early remnant of the solar nebula. So they fall back on a suggestion, put forward in the 1970s by Gel'mukhanov and Shalagin, that a different

process, which they call "light-induced drift", may do the trick instead.

In case that sounds like special pleading, a group from the University of Leiden has now, in a neat experiment, shown that even in relatively dense gases, molecular differentiation is possible (*Phys. Rev. Lett.* **70**, 742; 8 February 1993). Evidently H. I. Bloemink, J. M. Boom-Engering, E. R. Eliel and L. J. F. Hermans have it in mind that the segregation of hydrogen and deuterium takes place at a very early stage in the condensation of the Sun, when the temperature of the outer nebula is 800 kK or so, and molecules rather than atoms would predominate.

For light-induced drift to be effective, it is essential that the chance that a photon will be absorbed should depend on the motion of a molecule, towards or away from the source of radiation, and that absorption should then bias the kinematics of the molecules towards drift. Bloemink *et al.* have worked with water molecules (certainly constituents of the proto-nebula), irradiating them by a laser tuned to the frequency of the principal vibration band of the molecule, yet slightly to the red side of it while still within the Doppler-broadened width.

The guiding principle is that photons should be absorbed only by molecules moving towards the radiation source. The expectation that they would then, by virtue of their vibration, be more likely to collide with others and be randomized in velocity, leads directly to the idea that there should be a bias in their overall direction away from the source. The awesome thing is that exactly this phenomenon has been confirmed by irradiating capillary tubes 30 cm long containing water molecules and buffer gas (helium, for example). Water molecule concentrations have been increased by 2 or 3 parts of a million down the laser beam.

All that is a technical triumph, but two questions arise for the Solar System modellers. How can such a mechanism separate deuterium from hydrogen? Because the water molecule absorption bands are stronger than those of, say, HDO. And where in astrophysics is there a source of radiation fine-tuned as an expensive laser to be just to the red of the principal vibrational absorption band of H₂O? Why not the red-shifted output of vibrational photons from water molecules on the surface of the massive proto-Sun?

Bloemink *et al.* say that the numbers fit, and that light-induced drift would suffice to have cleaned the inner Solar System out to Venus of H₂O in about 10 million years. It sounds good.

John Maddox

Seasoned travellers

Carl D. Murray

THE obliquity of a planet, the angle that its spin axis makes with the perpendicular to its orbital plane, determines seasonal variations, and an understanding of its evolution is of fundamental importance to any study of climatic change. New research by Laskar and colleagues from the Bureau des Longitudes in Paris, reported on pages 608 and 615 of this issue^{1,2}, shows that the obliquities of all the terrestrial planets could have evolved chaotically at some time in their history. Although they show that the obliquity of Mars is still undergoing large, chaotic variations, the authors conclude that the Earth has been saved from a similar fate by the stabilizing influence of the Moon — the only large satellite in the inner Solar System.

The Earth has a modest obliquity of 23.5° but the resulting seasonal variations are sufficient to cause the annual migration of some of its inhabitants. The spin axis precesses in space at a current rate of about 50 seconds of arc per year and the obliquity varies by $\pm 1.3^\circ$ with a mean value of 23.3° (ref. 3). As the eccentricity of the Earth's orbit also changes in response to planetary perturbations, the combined effect over the millennia alters the annual amount of solar radiation reaching the surface of the Earth at a given latitude. According to the Milankovitch theory⁴, this is a driving mechanism for climatic change which may be associated with a succession of ice ages in the recent geological past.

In order to study the evolution of a planet's obliquity and precession, it is necessary to know how the orbit changes with time. In common with a number of other groups Laskar has investigated the prevalence of chaos in the Solar System by carrying out extensive numerical integrations of planetary orbits⁵. He knew how the orbits evolved and from a spectral analysis he had identified the dominant frequencies present in the orbital data for each planet. This is important, because if the rate of precession matches one of these frequencies a resonance can arise and large changes in the obliquity are possible.

Although the current spin rates of the planets are well determined, we can only speculate on their primordial values, at least as far as Mercury, Venus and the Earth are concerned. This is because they are all subject to strong tidal torques which affect each planet's spin. The slow retreat of the Moon from the Earth (3 cm yr⁻¹) and the consequent lengthening of the day are direct evidence of tidal evolution that implies that

the Moon was closer and the Earth spun faster in the past. Because the precession rate is proportional to the spin rate, resonances with the planetary perturbation frequencies could arise at various times.

By solving the precessional equations for a variety of initial obliquities and spin rates, Laskar and Robutel have now investigated the evolution of the obli-



Steadying influence: Earth and Moon pictured by NASA's space probe Galileo after it swung by last December on its way to Jupiter.

quity of Mercury, Venus, Earth and Mars¹. Given the ubiquitous nature of chaos in planetary motion^{5,6}, perhaps it is no longer surprising to discover that the obliquity behaves chaotically for a wide range of starting conditions for all these planets. In some ways, the possibility that this research could place constraints on the primordial rotation of the planets is a more interesting result. For example, the authors argue that Venus's obliquity could initially have been 0°, although its current rotation is retrograde, with an obliquity of 178° (making it unique among the planets).

Publication of this work conveniently coincides with a recent report⁷ which tackles the problem of the origin of primordial spin. In this, Dones and Tremaine argue that the spins of the terrestrial planets were determined by impacts with a few large bodies, left over at the end of the constructive phase of the Solar System. There are no obvious differences with the results of the French group; while Dones and Tremaine originally concluded that Venus's current rotation was the result of a late impact,

in a note added in proof they acknowledged the plausibility of the chaotic mechanism to change the obliquity.

Mars is a more interesting case. The results of Laskar and Robutel suggest that for a wide range of initial obliquities there is chaotic evolution with variations from 0–60° and a characteristic timescale for exponential divergence of five million years. It has been recognized for some time that Mars undergoes large amplitude changes in its obliquity and orbit⁸; the work of Laskar and Robutel simply demonstrates that the obliquity changes are larger than expected and unpredictable. The consequences for the martian climate must be profound and with the planet's geological record awaiting examination, the exciting prospect of studying the climatology of a second planet may soon be a reality.

With the Earth's obliquity it is a case of what might have been and what will be. In a further paper, Laskar, Joutel and Robutel² study the effect of the Moon on the Earth's obliquity. They discovered a chaotic zone for initial values of the obliquity in the range 60–90°. However, the Moon exerts a torque on the Earth and in its absence the chaotic zone would extend from 0° to 85° with values exceeding 50° for a few million years even for low initial obliquities. Given that variations as small as $\pm 1.3^\circ$ may trigger ice ages and that for obliquities greater than 54° the Equator receives less radiation annually than the poles⁸, the forecast for a Moon-less Earth would have been bleak.

Although the primordial spin rate of the Earth and the connected problem of the origin of the Moon are contentious issues, the future of the Earth–Moon system is more assured but not reassuring. Ward⁹ has calculated the fate of the system when a series of spin-orbit resonances are encountered as the Moon's orbit approaches about 68 Earth radii (it is currently about 60). Although his

1. Laskar, J. & Robutel, P. *Nature* **361**, 608–612 (1993).

2. Laskar, J., Joutel, F. & Robutel, P. *Nature* **361**, 615–617 (1993).

3. Laskar, J. et al. *Astr. Astrophys.* (in the press).

4. Imbrie, J. *Icarus* **50**, 408–422 (1982).

5. Laskar, J. *Icarus* **88**, 266–291 (1990).

6. Sussman, G. J. & Wisdom, J. *Science* **257**, 56–62 (1992).

7. Dones, L. & Tremaine, S. *Science* **259**, 350–354 (1993).

8. Ward, W. R. in *Mars* (eds Kleffler, H. H. et al.) 298–320 (University of Arizona Press, Tucson, 1992).

9. Ward, W. R. *Icarus* **50**, 444–448 (1982).

work may need revising in light of Laskar's results, he predicts an obliquity near 60° some two billion years in the future.

Ward began a recent review article⁸ on the dynamics of Mars with the sentence, "No object in the solar system is dynamically invariant." Given the current rate of discoveries in this

field, perhaps the next challenge for Solar System dynamicists will be to find objects that *never* exhibit unusual dynamical behaviour. □

Carl D. Murray is in the Astronomy Unit, Queen Mary and Westfield College, University of London, Mile End Road, London E1 4NS, UK.

SOLID-STATE PHYSICS

Current oscillations ring true

Sean Washburn

THE subtle quantum-mechanical interplay between current and magnetic field gives rise to current oscillations in devices such as SQUIDS (superconducting quantum interference devices), which are small loops of superconducting material. On page 617 of this issue¹, Moschalkov *et al.* describe a superconducting loop, far smaller than a typical SQUID, which gives rise to similar oscillations, although working by a different mechanism. Oscillations are also seen in a new silver interferometer loop, to which superconducting 'mirrors' have been added, reported by Petrashov *et al.*². In each case, the long-term prospect is of electronic circuit elements based on these oscillations.

The oscillations arise because, at low temperature, even normal conducting electrons retain the memory of their quantum-mechanical wavefunction over distances of up to several micrometres³ (at higher temperatures, thermal effects scramble the wavefunction's phase). Magnetic flux threaded through the middle of a loop no larger than this alters the phase of electrons passing along either side, through the Aharonov-Bohm effect, so that current passing in through one contact and out through a second is modulated by interference between the phase-shifted wavefunctions.

Seeing the coherence effect, however, requires temperatures lower than 1 kelvin and extremely low excitation levels (voltage below 0.1 millivolts) to avoid smearing out the delicate quantum coherence. Another barrier to using this effect in 'transistors' is that the quantum coherence occurs between particular pairs of electron trajectories under particular circumstances such as the value of magnetic field, the voltage bias across the sample, the starting temperature of the electrons, and the sample configuration — including the configuration of impurities in it. Any change in the circumstances leads to a new interference pattern. In fact, in many experiments the interference pattern is sensitive to moving so much as a single impurity in the device.

In contrast, the response of a superconducting circuit, which relies on the collective motion of the superconducting Cooper pairs of electrons, is relatively impervious to disorder and to the details of the sample configuration. The beauty

of the two new experiments is that they combine features of the conventional SQUID with the mesoscopic effects in very small interferometers.

The experiment by Moschalkov *et al.* uses a small simple superconducting loop (Fig. 1), which is equivalent to the configurations used to study flux quantization (flux trapping) for decades⁴. The magnetic field through a superconducting loop can only be a multiple of the discrete flux unit $\Phi_0 = h/2e$ (where h is Planck's constant, and e is the electron charge). If the ambient field is fractionally larger, the difference is expelled from the loop by screening currents. Only when the ambient flux increases by $\Phi_0/2$ does the screening current weaken enough so that a flux quantum pops into the loop, and then the screening current must reverse direction to keep the extra flux (the difference between the trapped

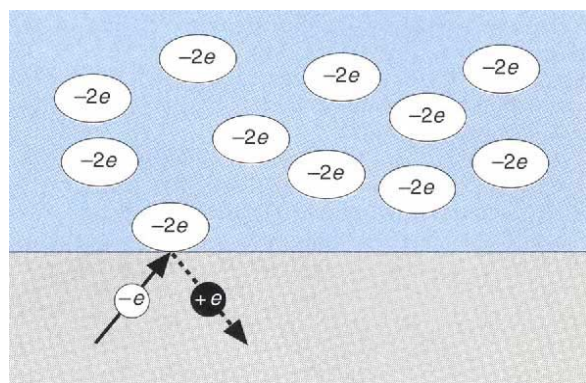


FIG. 2 Andre'v reflection: an electron incident on a superconducting electrode (blue) can enter only by creating a Cooper pair. The second electron of the pair has to be 'stolen' from the normal material, so creating a 'hole' which acts as a positive charge reflected from the interface. The net effect is to double the number of carriers.

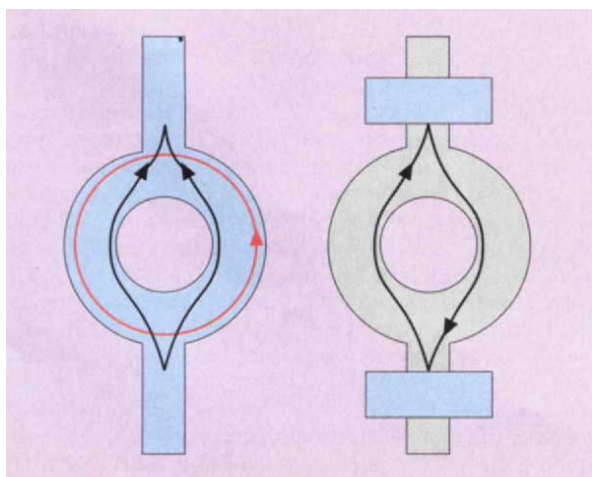


FIG. 1 Two ways of using mesoscopic loops (less than $1 \mu\text{m}$ in diameter) to give clear current oscillations. Left, with superconducting material, the current passes either side of a hole threaded with flux, which shifts their relative phase; if the ambient field raises the screening current (red) high enough on one side, that side will turn normal, allowing additional flux to enter the ring¹. Right, superconducting mirrors (blue) either side of a metallic ring cause Andre'v reflections which amplify and stabilize field-induced oscillations².

flux and the ambient) from escaping back outside.

In a SQUID, an artificial weak link (a tunnel barrier) is inserted into the loop to give the flux a place to pass more easily into the centre; the screening current can be used to detect very small changes in the ambient magnetic field. In fact, the most sensitive detectors known are SQUIDS (capable of sensing changes smaller than $10^{-6}\Phi_0$) having noise levels within an order of magnitude of the limit set by Heisenberg's uncertainty principle⁵.

The new wrinkle is that Moschalkov and colleagues' loop (without a tunnel barrier) is so small that the Cooper-pair coherence length is the size of the loop and the screening currents themselves provide a weak link: on one side the screening current adds to the supercurrent and pushes it closer to the critical current beyond which the superconductivity breaks down and the loop goes normal. The results of this experiment

are nicely in line with theoretical predictions for small superconducting loops⁶.

Petrashov *et al.*, in contrast, use strips of superconductor (either aluminium or a lead-gold alloy) as mirrors either side of a silver interferometer loop, to reflect the conducting electrons and so control the oscillation pattern. The 'Andre'ev' reflection occurs when a normal electron changes to a Cooper pair as it crosses into the superconductor (Fig. 2). To conserve charge, a hole (a collective excitation with charge $-e$) is reflected back along the electron's path. In the simplest view this doubles the number of interfering carriers⁷, but the experimenters discovered a much larger (more than a factor of 100) increase in the oscillation amplitude. The results of the experiment are in spectacular disagreement with theoretical calculations. Even more surprising, the mirrors appear to make the electrons 'more coherent', so that the Aharonov-Bohm oscillations survive at much higher temperatures. Apparently, the study of the coupling of superconducting and normal materials is rich with surprises.

As already mentioned, SQUIDS are powerful tools without parallel when it comes to measuring minute electrical or magnetic signals. Can resistance oscillations from quantum coherence also be employed in useful circuits? The two new circuits point to a new way to control the debilitating randomness in the response of disordered materials (and for the near future, at least, we will be forced to accommodate disorder in all detectors). The papers, that by Petrashov *et al.* in particular, suggest that mesoscopic effects may eventually be useful.

On the other hand, can such devices compete with technologically less demanding circuits? Manufacturing mesoscopic circuits is a formidable task, and the present devices need to be cooled even lower than conventional SQUIDS to operate successfully. So for the time being, at least, the conventional SQUID, with its quantum-limited noise floor, will remain the tool of choice. □

Sean Washburn is in the Department of Physics and Astronomy, University of North Carolina, Chapel Hill, North Carolina 27599, USA.

Playing chess with reverse transcriptase

Douglas D. Richman

FOR an organism to survive a toxic challenge, resistant mutants must evolve, and it is the agility of pathogens in generating such mutations that prompts the continual search for new insecticides, drugs for cancer chemotherapy and antimicrobial agents. On page 650 of this issue, Chow *et al.*¹ put forward a strategy to turn this evolutionary survival mechanism against the human immunodeficiency virus (HIV): by selecting certain combinations of mutation for drug resistance, they propose, it may be possible to force the virus into inviability.

Chemotherapy for HIV infection prolongs the disease-free interval in asymptomatic patients, improves the quality of life in symptomatic patients and delays death. The deoxynucleoside analogues used — AZT (zidovudine, 3'-azido, 3'-deoxythymidine), ddC (zalcitabine, 2', 3'-dideoxycytidine) and ddI (didanosine, 2', 3'-dideoxyinosine) — work by inhibiting the viral enzyme, reverse transcriptase. Any satisfaction generated by these accomplishments, however, must be set against the knowledge that nucleoside chemotherapy only buys time. Levels of virus replication (as measured by virus infectivity, serum levels of HIV p24 antigen, or viral RNA assayed by the polymerase chain reaction in plasma or serum) are only diminished, not completely suppressed by these drugs. Disease progresses despite therapy, which could be due to incomplete suppression of virus, as well as to the emergence of viral mutants with reduced susceptibility to these nucleoside drugs^{2,3}. For a number of reasons it has been difficult to document the proportion of drug failure attributable to the emergence of this nucleoside resistance (for review see ref. 3).

A second class of promising inhibitors of HIV-1 replication also inhibits reverse transcriptase (RT). Numerous, chemically distinct non-nucleoside compounds share a number of attributes including high potency, low toxicity, excellent pharmacokinetic properties and synergy with nucleoside agents. But the rapid selection *in vitro* of highly resistant mutants of HIV suggested that, as useful drugs, these 'non-nucleoside reverse transcriptase (NNRT) inhibitors might possess an Achilles heel'^{4,5}. In initial clinical trials the drugs quickly reduced measurements of virus replication and they increased CD4 lymphocyte counts; however, these promising activities dissipated in about one month, at about the same time as drug-resistant mutants be-

gan to appear^{6,7}. These observations are the most compelling evidence that the emergence of drug resistance can result in drug failure.

Mutations in the RT of HIV-1 that account for reduced susceptibility to drugs have been identified by sequencing isolates of resistant virus, and confirming the contribution of identified mutations to the reduction of susceptibility by using site-directed mutagenesis to place these mutations in an infectious plasmid containing a sensitive HIV genome. Mutations conferring reduced susceptibility to each of the RT inhibitors have been identified in this way. The mutations conferring resistance to the NNRT inhibitors occur in amino-acid residues 100 to 108 and 181 to 190, which comprise two β -sheets that form the binding pocket for these compounds⁸. The X-ray crystallographic structure of the RT of HIV-1 was made possible by stabilizing crystals of the enzyme with nevirapine, one of the NNRT inhibitors⁸. The binding site of these drugs is separate from but adjacent to the presumed catalytic and nucleotide-binding sites.

Several interesting interactions between drug-resistance mutations in the RT have been reported. The first AZT mutation to appear (amino-acid residue 70) usually disappears, even under the selective pressure of AZT therapy, when the more potent mutation at residue 215 appears, probably because they are antagonistic⁹. Virus with both mutations is less resistant than virus with the mutation at 215 alone. AZT mutations at residues 41 and 215 are synergistic, conferring 64-fold reductions in susceptibility together but 4 and 16-fold reductions separately¹⁰. The ddI resistance mutation at residue 74 and the NNRT-resistance mutation at residue 181 have no effect on AZT susceptibility in wild-type virus, but confer a compensatory reduction in AZT resistance in virus containing the 215 mutation^{11,12}.

Chow *et al.*¹ now offer some data to support a strategy to exploit these mutational interactions. A successful mutation for drug resistance can be defined as one that permits the organism to replicate normally despite the presence of the inhibitor. For example, the tyrosine to cysteine mutation at residue 181 results in an RT that functions like the wild-type enzyme but has a K_i more than 100-fold higher than the NNRT inhibitors^{4,5}. Chow *et al.* demonstrate that successful mutations for resistance to a nucleoside and to a NNRT inhibitor

1. Moschalkov, V. V., Gielen, L., Dhallé, M., van Haesendonck, C. & Bruynseraede, Y. *Nature* **361**, 617–620 (1993).
2. Petrashov, V. T., Antonov, V. N., Delsing, P. & Claesson, R. *Phys. Rev. Lett.* **70**, 347–350 (1993).
3. Washburn, S. & Webb, R. A. *Rep. Prog. Phys.* **55**, 1311–1383 (1992).
4. Deaver, B. S. & Fairbank, W. F. *Phys. Rev. Lett.* **7**, 43–46 (1961).
5. Ketchen, M. B. & Jaycox, J. M. *Appl. Phys. Lett.* **40**, 736–738 (1982).
6. Fink, H. J., Lopez, A. & Maynard, R. *Phys. Rev.* **B26**, 5237–5240 (1982).
7. Spivak, B. Z. & Khmelinskii, D. E. *JETP Lett.* **35**, 412–416 (1982).

may be incompatible with each other — that is, two different drugs in combination may select for mutations that produce dysfunctional enzyme and poorly replicating or non-replicating virus.

Effective combination chemotherapy, for example for tuberculosis or malignancies, typically combines drugs that target different functions. Drug synergy has been accomplished when the second drug acts at a sequential step in metabolism (for instance trimethoprim and sulfa in folate synthesis) or amplifies the activity of the other (for instance penicillin inhibits bacterial cell-wall synthesis to facilitate entry by aminoglycosides). So it is intriguing that two different inhibitors of RT, the nucleoside AZT and the NNRT nevirapine, are potentially synergistic¹³.

The mutations induced by these different classes of drug each produce a functional but resistant virus; however, the combination of resistance mutations may disable the virus or attenuate its activity. Chow *et al.* thus propose a strategy of combination chemotherapy with drugs targeted at the same enzyme. This strategy, which they call convergent combination therapy, resembles a chess game in which several RT inhibitors are positioned to force the RT to make a fatal combination of mutations. HIV is a highly mutable and wily opponent, however, and that it may be able to avoid checkmate with a large array of alternative moves (mutations) is a real possibility. Both *in vitro*¹² and clinical data (unpublished observations) suggest that this may be the case. Nevertheless, the progress of the HIV epidemic and the limited number of therapeutic alternatives at our disposal mean that the strategy of convergent combination therapy has to be investigated. Several clinical trials to evaluate the approach are in progress, and several more are in the planning stage. □

Douglas D. Richman is in the Departments of Pathology and Medicine, University of California San Diego, La Jolla, California 92093-0679, USA.

The new pulsar next door

Andrew Fruchter

MILLISECOND pulsars, neutron stars rotating at rates not far short of 1 kilohertz, were thought until recently to be comparatively rare. So it comes as a welcome confirmation of current ideas to learn, on page 613 of this issue¹, that one lurks virtually on our galactic doorstep, just 400 light-years away. The discovery of this hitherto unknown neighbour was made by Johnston *et al.* using the Parkes radio telescope in Australia, during a systematic survey of the sky for pulsars.

Astronomers tend to view statistical arguments with suspicion, and conclusions based upon small numbers are met with something approaching derision. Therefore, when Kulkarni and Narayan² used the ages and distances of the three then-known millisecond pulsars to derive a galactic birthrate and population density, their ingenuity was praised, but their results discounted, for their conclusions directly challenged the standard theory of millisecond pulsar creation³. Millisecond pulsars are not born to great rotation rates, but are made in stellar binaries. This happens when a neutron star comes close enough to its binary partner (through orbital decay, or because the companion inflates with old age) to strip off its neighbour's outer atmosphere. The angular momentum of the matter that rains down on the neutron star will spin it up, in some cases to rotation rates of hundreds of times per second (the two fastest known millisecond pulsars spin more than 600 times per second). Owing to the strength of the gravitational field surrounding a neutron star, the accreting matter is heated to X-ray temperatures, and the embryonic millisecond pulsar appears as a low-mass X-ray binary.

Kulkarni and Narayan used the results of a number of searches for millisecond pulsars to estimate the birthrate of these stars and compared their result with theories of the evolution of low-mass X-ray binaries. They found that millisecond pulsars are born far more frequently than one would predict from theory and the observed numbers of low-mass X-ray binaries (and so a near-by example would not be so unlikely).

The key to this calculation was an accurate estimate of the population density of millisecond pulsars. Low-mass X-ray binaries are extremely bright, and it is thought that most of the active members of this population in our Galaxy have been catalogued⁴. Their number density, then, is known. Furthermore, the age of a pulsar can be estimated by dividing its period by twice

the rate at which the period is observed to lengthen with time. (The factor of two depends upon the physics by which pulsars radiate away their energy, which is only partially understood; however, factors of two are not crucial to the argument.) Given a population density then, pulsar ages allow one to obtain a birthrate.

The galactic millisecond-pulsar population that Kulkarni and Narayan derived and which has been supported by later, more sensitive pulsar surveys^{5,6} exceeded 100,000. This is roughly equal to the number of slow normal pulsars thought to be active in the Galaxy. So it should perhaps not be too surprising that the newly discovered millisecond pulsar now vies for the title of the nearest known neutron star. (One of its competitors, Geminga, has also only just been identified as a pulsar⁷; until last year it was known only as another one of many astronomical puzzles.)

It is easiest to find the distance to a pulsar if it is associated with some other object — a globular cluster or a supernova remnant — whose distance is readily measured. In the absence of such aids, astronomers must generally settle for an estimate based upon the dispersion of the radio pulses as they traverse the ionized interstellar medium. The interstellar medium, however, is notoriously inhomogeneous, and such estimates are prone to large errors. Nonetheless, for the new pulsar, PSR J0437-4715, that is the best that can be done immediately. However, if this pulsar is at a distance at all comparable with the estimated 400 light-years, it should soon be possible to measure its parallax through pulse timing. Over the course of a year, the time of arrival of the neutron star's pulses should vary by about 10 μ s, owing to the Earth's motion and the spherical wavefront of the pulsar's radiation. This effect could be measured with the 15 μ s accuracy of the reported timing. But with the pulsar's narrow radio profile and the extraordinary strength of its signal, simple technical improvements should make possible arrival times accurate to better than 1 μ s.

A measurement of parallax will allow astronomers to invert the usual procedure of determining a pulsar's distance: from the dispersion measure they could measure the local interstellar electron density. Furthermore, pulse timing will quickly reveal the pulsar's motion across the sky. Its velocity along our line of sight might be determined if the optical counterpart proposed by another group of astronomers⁸ is correct: spec-

1. Chow, Y.-K. *et al.* *Nature* **361**, 650–654 (1993).
2. Larder, B. A., Darby, G. & Richman, D. D. *Science* **243**, 1731–1734 (1989).
3. Richman, D. D. *Antimicrob. Ag. Chemother.* (in the press).
4. Nunberg, J. H. *et al.* *J. Virol.* **65**, 4887–4892 (1991).
5. Richman, D. D. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **88**, 11241–11245 (1991).
6. Kappes, J. C. *et al.* (abstr.) *VIII Int. Conf. AIDS*, Amsterdam, 19–24 July, 1992.
7. Richman, D. D. (abstr.) *VIII Int. Conf. AIDS*, Amsterdam, 19–24 July, 1992.
8. Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A. & Steitz, T. A. *Science* **256**, 1783–1790 (1992).
9. Boucher, C. A. B. *et al.* *J. infect. Dis.* **165**, 105–110 (1992).
10. Kellam, P., Boucher, C. A. & Larder, B. A. *Proc. natn. Acad. Sci. U.S.A.* **89**, 1934–1938 (1992).
11. St Clair, M. H. *et al.* *Science* **253**, 1557 (1991).
12. Larder, B. A. *Antimicrob. Ag. Chemother.* **36**, 2664–2669 (1992).
13. Richman, D. D. *et al.* *Antimicrob. Ag. Chemother.* **35**, 305–308 (1991).

troscopy could reveal the object's Doppler shift. Variations in the dispersion as the pulsar moves would then map out inhomogeneities in the interstellar medium.

For many pulsar observers, however, the most remarkable aspect of this discovery may be the ease with which one can detect individual pulses. The strength of this pulsar's signal will allow studies of the microstructure and polarization properties at a level so far possible only on slower, generally brighter pulsars. Although the past few years have brought a greater understanding of the phenomenology of emission from slow pulsars⁹, millisecond pulsar emission, and much of emission physics for that matter, remain poorly understood. The detection of X-rays by Rosat⁸ is therefore important. Observations by the Compton satellite are likely to be even more crucial; if current theories are correct, this pulsar should be a bright γ -ray source.

The search that discovered PSR J0437-4715 has now covered about one sixth of the celestial sphere (R. Manchester, personal communication). As another sixth of the sky had previously been searched for fast pulsars, this object may remain the nearest known millisecond pulsar. As neutron stars go, it is a rather unimpressive radio transmitter. Proximity is the cause of its apparent brightness. Continuing pulsar searches are thus likely to discover a number of comparably bright but perhaps somewhat more distant millisecond pulsars. This prediction should please cosmologists nearly as much as radio astronomers, for gravitational waves created in the early Universe may perturb the apparent distances to stars at a level that is detectable through the cross-comparison of millisecond-pulsar timing residuals^{7,8}. In the foreseeable future, the nearest of pulsars may provide a probe of the most distant Universe. □

Andrew Fruchter is in the Department of Astronomy, University of California, Berkeley, California 94720, USA.

Promiscuous liaisons

Stephen Green

ON page 657 of this issue¹, Carlberg and colleagues describe experiments which show that the receptor for vitamin D₃, known as VDR, can act through two different signalling pathways to activate gene expression. This finding has several implications for the field concerned, the long-term object in which is to understand how complex biological processes are regulated by simple molecules such as steroids, thyroid hormones, retinoids and vitamin D₃ itself.

These molecules interact specifically with members of a large superfamily of ligand-activated transcription factors termed the nuclear hormone receptor superfamily^{2,3}. More than 30 members of the family have been identified, including a number of 'orphan' receptors still in search of ligands⁴. Studies of the oestrogen and glucocorticoid receptors indicated that the classical steroid hormone receptors form homodimers and recognize short palindromic DNA sequences frequently located upstream of their target genes^{5,6}. However, the simple concept that each hormone binds to a specific receptor homodimer acting through a unique palindromic DNA sequence is fast becoming the exception rather than the rule, as exemplified in the report by Carlberg *et al.*

The mechanism of retinoid action differs in a number of ways from that of steroid hormone action. Six members of the receptor superfamily are activated by retinoic acid which explains, in part, the complex biology of retinoid action. The three retinoic acid receptors (RAR α , β , γ and a number of isoforms) bind all-trans retinoic acid⁷, whereas the three retinoid X receptors (RXR α , β and γ) bind the isomer 9-cis-retinoic acid^{8,9}. Although each receptor class is highly conserved within the ligand and DNA-binding domains, there is surprisingly greater homology between the ligand-binding domain of RAR and a thyroid hormone receptor (TR α or β) than there is between RAR and RXR.

One piece of evidence indicating that VDR, RAR and TR behave in a different way to the steroid hormone receptors is that they require a nuclear factor to bind DNA *in vitro*. The dramatic finding, however, was that RAR, TR and VDR all use the same nuclear factor, and that this is RXR¹⁰⁻¹³. Unlike steroid hormones, the natural response elements for retinoic acid, thyroid hormones and vitamin D₃ consist of a direct repeat (DR) of a binding site (related to the hexamer AGGTCA) rather than a palindrome. VDR, RAR and TR bind to their response elements as heterodimers

with RXR and the spacing of the direct repeat determines the binding preference — 3 base pairs (bp) spacing for VDR:RXR; 4 bp for TR:RXR; and 5 bp for RAR:RXR (see figure)^{10,14}. This binding preference led to the proposal of the so-called 3-4-5 rule, according to which each heterodimer acts through a unique direct repeat¹⁴.

Carlberg *et al.* have now come up with support for these earlier findings, using *in vitro* DNA-binding assays and transfection experiments to express VDR and RXR in *Drosophila* cells. They demonstrate that VDR forms a heterodimer with RXR and acts through the mouse osteopontin vitamin D₃ response element that contains a direct repeat spaced by 3 base pairs (DR3). They also show that VDR can act without RXR to activate the human osteocalcin vitamin D₃ response element. This element contains three directly repeated binding sites, two separated by 6 bp (DR6) and a third imperfect repeat separated by 3 bp (DR3), thus suggesting degeneracy within the 3-4-5 rule. In this case, VDR may bind as a homodimer or by interaction with another factor such as the *Drosophila* nuclear receptor ultraspiracle¹⁵. These findings are likely to be biologically relevant, as Carlberg *et al.* observed the same results using endogenous VDR and RXR in mammalian cells.

The 3-4-5 rule has also been challenged by the observation that RAR/RXR heterodimers cooperate to activate the genes that contain retinoic acid response elements with 1 bp (DR1) and 2 bp (DR2) spacings^{7,16}. Furthermore, a COUP-TF (chicken ovalbumin upstream promoter transcription factor) homodimer recognizes a number of differently spaced direct repeat motifs and appears to act as a repressor¹⁷. One possible explanation of these findings is that the various receptor homo- and heterodimers interact with the different direct repeats with different affinities, and that DNA binding is not sufficient for transcriptional activity. For example, the transcriptional activity would be influenced by other transcription factors that bind to neighbouring response elements.

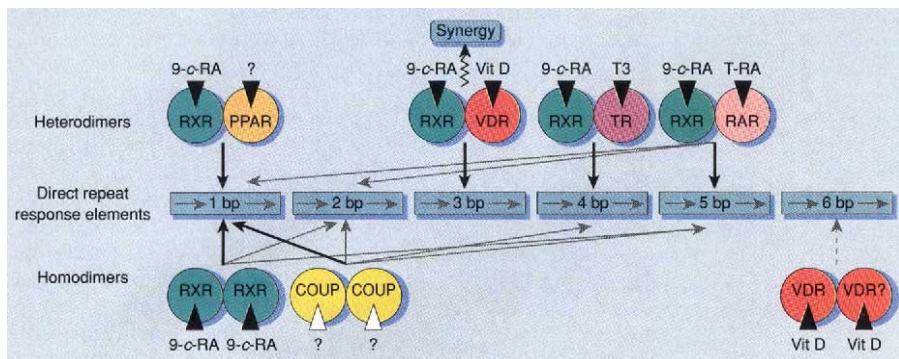
Overall, the findings of the past year have raised some intriguing biological questions. Why should one receptor, RXR, interact with many others including several orphan receptors? And what, if any, is the role of its ligand 9-cis-retinoic acid? Unlike receptors for steroid hormones, those for thyroid hormones, retinoids and vitamin D₃ do not seem to require the ligand in order

1. Johnston, S. *et al.* *Nature* **361**, 613-615 (1993).
2. Kulkarni, S. R. & Narayan, R. *Astrophys. J.* **335**, 755-768 (1988).
3. Bhattacharya, D. & van den Heuvel, E. P. J. *Phys. Rep.* **203**, 1-124 (1991).
4. McClintock, J. & Rappaport, S. A. in *Cataclysmic Variables and Low-Mass X-ray Binaries* 61 (Reidel, Dordrecht, 1985).
5. Fruchter, A. S. thesis, Univ. Princeton (1989).
6. Johnston, S. & Bailes, M. *Mon. Not. R. astr. Soc.* **252**, 277-281 (1991).
7. Halpern, J. P. & Holt, S. S. *Nature* **357**, 222 (1992).
8. Becker, W., Trumper, J., Brazier, K. T. S. & Belloni, T. *IAU Circ. No.* 5701 (1993).
9. Rankin, J. M. *Astrophys. J.* **352**, 247-257 (1990).
10. Romani, R. W. in *Timing Neutron Stars* (eds Ögelman, H. & van den Heuvel, E. P. J.) 113-117 (Kluwer, Dordrecht, 1989).
11. Thorne, K. S. in *Recent Advances in General Relativity* (eds Janis, A. & Porter, J.) 196-229 (Birkhauser, Boston, 1992).

to bind tightly to DNA. One action of 9-*cis*-retinoic acid is to promote the formation of RXR homodimers that bind to a number of retinoid response elements¹⁸, although none appears to be specific for RXR homodimers. In the case of heterodimers, RXR may act simply as an adapter to allow other receptors to bind specifically to their response elements. However, in addition, RXR could contribute to the trans-

response. A possible function of such a mechanism would be to control the switch from (for example) cell proliferation to differentiation, with the hormone regulating a different or additional gene network in each case.

The real picture is no doubt more complicated than that uncovered to date. For example, several orphan receptors (and probably others that remain to be discovered) also form heterodimers with



Promiscuous liaisons among the nuclear hormone receptor family. The various interactions are discussed in the text. RXR, receptor for 9-*cis*-retinoic acid (9-*c-RA*); PPAR, receptor activated by peroxisome proliferators and fatty acids⁴; VDR, vitamin D receptor; TR, thyroid hormone receptor; RAR, receptor for *trans*-retinoic acid; COUP, chicken ovalbumin upstream promoter.

criptional activity of the heterodimer and this property could be regulated by 9-*cis*-retinoic acid. The new findings of Carlberg *et al.*¹, and those of others^{16,19}, support this second view. In the presence of both VDR and RXR, vitamin D₃ and 9-*cis*-retinoic acid act synergistically on the osteopontin response element (DR3), whereas RXR or its ligand has no effect on the osteocalcin response element (DR6). Presumably, the ability of RXR to contribute to transcriptional activation of the heterodimer will depend upon the context of the response element within the promoter, and also the cell type, thus providing further fine tuning of the response^{16,20}.

The ability of several members of the nuclear hormone receptor superfamily to form heterodimers with RXR increases the combination of transcription factors that can have unique biological properties. However, this could mean that 9-*cis*-retinoic acid is a general permissive factor that acts synergistically with other ligands such as vitamin D₃, all-*trans* retinoic acid and thyroid hormones to amplify the hormonal signal. So a cell that already contains 9-*cis*-retinoic acid would be exquisitely sensitive to the presence of the second hormone. That hormones such as vitamin D₃ can work in two ways also provides a mechanism for regulating two gene networks. In the absence of 9-*cis*-retinoic acid both networks would respond similarly. However, in its presence, one network would be more sensitive to the second hormone and would achieve a greater maximal

RXR and the biological consequence of many receptors each competing for the attention of RXR is uncertain. In addition, other heterodimer combinations have been reported, for example between RAR and TR²¹, and some receptors interact weakly with their response elements as homodimers and monomers, suggesting that other signalling pathways exist²².

Stephen Green is at ZENECA Central Toxicology Laboratory, Alderley Park, Macclesfield, Cheshire SK10 4TJ, UK.

1. Carlberg, C. *et al.* *Nature* **361**, 657–660 (1993).
2. Evans, R. M. *Science* **240**, 889–895 (1988).
3. O'Malley, B. W. & Conneely, O. M. *Molec. Endocr.* **6**, 1359–1361 (1992).
4. Issemann, I. & Green, S. *Nature* **347**, 645–650 (1990).
5. Green, S. & Chambon, P. *Trends Genet.* **4**, 309–314 (1988).
6. Luisi, B. F. *et al.* *Nature* **352**, 497–505 (1991).
7. Leid, M., Kastner, P. & Chambon, P. *Trends Biochem.* **17**, 427–433 (1992).
8. Levin, A. A. *et al.* *Nature* **355**, 359–361 (1992).
9. Heyman, R. A. *et al.* *Cell* **68**, 397–406 (1992).
10. Yu, V. C. *et al.* *Cell* **67**, 1251–1266 (1991).
11. Leid, M. *et al.* *Cell* **68**, 377–395 (1992).
12. Zhang, X. K., Hoffmann, B., Tran, P. B. V., Graupner, G. & Pfahl, M. *Nature* **355**, 441–446 (1992).
13. Kliewer, S. A., Umesono, K., Mangelsdorf, D. T. & Evans, R. M. *Nature* **355**, 446–449 (1992).
14. Umesono, K. *et al.* *Cell* **65**, 1255–1266 (1991).
15. Yao, T. P., Seagraves, W. A., Oro, A. E., McKeown, M. & Evans, R. M. *Cell* **71**, 63–72 (1992).
16. Durand, B., Saunders, M., Leroy, P., Leid, M. & Chambon, P. *Cell* **71**, 73–85 (1992).
17. Cooney, A. J., Tsai, S. Y., O'Malley, B. W. & Tsai, M. J. *Molec. cell. Biol.* **12**, 4153–4163 (1992).
18. Zhang, X. K. *et al.* *Nature* **358**, 587–591 (1992).
19. Kliewer, S. A., Umesono, K., Noonan, D. J., Heyman, R. A. & Evans, R. M. *Nature* **358**, 771–774 (1992).
20. Nagpal, S. *et al.* *Cell* **70**, 1007–1019 (1992).
21. Glass, C. K. *et al.* *Cell* **59**, 697–708 (1989).
22. Forman, B. M., Casanova, J., Raaka, B. M., Ghysdael, J. & Samuels, H. H. *Molec. Endocr.* **6**, 429–442 (1992).

RÉSUMÉ

Stepping out

SKATES and rays — a suborder (Rajoidea) of batoid fish — didn't take part in *The Lobster Quadrille*. If Lewis Carroll had known of the remarkable capacity of certain species for bipedal locomotion, however, he might well have invited them to join the dance. There have been allusions to such a capacity. But R. J. Holst and Q. Bone (*Phil. Trans. R. Soc. B339*, 105–108; 1993) now show by dissection of the pelvic fins and ciné photography of the fish in action that they can and do use limb-like parts of the fins for movement akin to walking. The authors include photos of the resulting tracks made in soft mud — they suggest that rajids 'walk' when feeding on the bottom, for that form of motion enables their invertebrate prey to be located and captured most easily, and that their tracks may be evident in the fossil record.

Water pump

EXPERIMENTS at close to absolute zero have shown how light can push atoms around, but it seems that this has all been done before — five billion years ago when the Sun started to roast its protoplanetary nebula. H. I. Bloemink *et al.* have irradiated a tube of water vapour in a background of hydrogen with infrared radiation, and find that a concentration gradient builds up as the radiation presses harder on the H₂O than on the background gas (*Phys. Rev. Lett.* **70**, 742–745; 1993). Their aim was to explain the gradient of the deuterium/hydrogen ratio in the Solar System: the concentration ratio is 10⁻² in Venus's atmosphere but 10⁻⁵ in Jupiter's. The hot, young Sun may have driven H₂O away, the authors suggest, but not HDO, which responds less strongly to light. It remains to be seen whether planetary scientists will be swayed by the argument, given that more brute forces operated later in planetary atmospheres.

More myosin

PHYLOGENETIC analysis of the force-generating protein myosin reveals that there are at least three classes of the protein in eukaryotes (H. V. Goodson and J. A. Spudich, *Proc. natn. Acad. Sci. U.S.A.* **90**, 659–663; 1993). Myosin had been divided into two classes, I (single-headed) and II (the better known two-headed version). But the new analysis of amino-acid sequences shows that myosin I consists of a 'classic' type and a 'dilute' type, as divergent from each other as each is from myosin II. Further investigation may demonstrate that as many as seven classes of myosin I have evolved separately. In the myosin II class, Goodson and Spudich find that smooth muscle and striated muscle myosin were independently derived from non-muscle myosin.

Oceans of change

Lee Kump

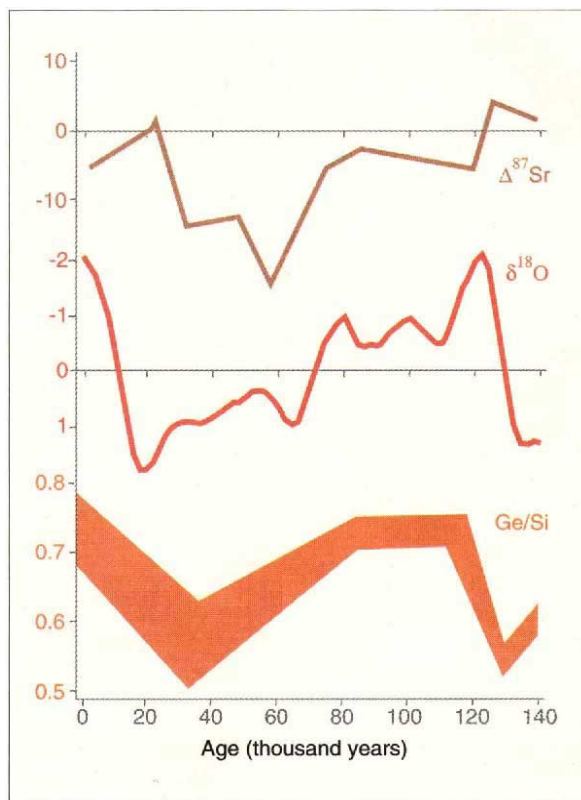
CLIMATE change is not just a matter of average global temperatures, although those are talked about most. Regional changes in temperature and water balance also play an important role, and through erosional processes these affect how changes in climate over geological timescales have been recorded in sedimentary rocks. Earlier last year, Dia *et al.*¹ found that strontium isotope ratios of sediments deposited on the sea floor have recorded changes in the nature of chemical weathering on the continents during the ice ages. And now Froelich *et al.*² find that sedimentary germanium/silicon ratios signal the same. But curiously, these two newly discovered indicators lead to seemingly paradoxical conclusions about climate change from glacial to interglacial conditions.

To infer changes in chemical weathering from these indicators requires an understanding of the controls of ocean chemistry, because in both cases it is the ocean's composition that is actually being recorded in the sediments. In the past, dogma had it that the oceans were controlled by equilibrium with the sediments on the sea floor, suggesting little change in composition even on geologically long timescales. More recently a new paradigm has arisen, one that states that the oceans are in a state of flux, subject to a dynamic balance between inputs and outputs.

Now, however, even this picture, which still hints at stability over long timescales, is being called into question. Over the past year, high-resolution studies have been reported showing long-term trends in ocean composition. These suggest that the rates of delivery of dissolved material, as well as its composition, have varied in concert with the waxing and waning of the continental glaciers. As a result, the chemical composition of sea water has undergone measurable variations on the 100,000-year timescales of Quaternary glaciation. Best studied of these are changes in the oxygen isotope composition of sea water, reflecting changes in temperature and continental ice volume. The two new records, strontium isotope ratios¹ and germanium/silicon ratios^{2,3} appear to be a much more direct measure of weathering itself.

If these records resulted entirely from variations in riverine fluxes (the largest source of dissolved material to the oceans), substantial changes to those fluxes are indicated. To generate the low strontium isotope ratios observed in sediments from the last glacial period (50,000 years ago), a 50 per cent reduc-

tion in the rate of transfer of strontium from the land to the sea is required¹. Somewhat troubling is that the Ge/Si ratio shows the opposite change — a doubling of the silica flux (and a 70 per cent increase in the germanium flux) — if one accepts the inverse relationship proposed by Froelich *et al.*² between the Ge/Si ratio and silica content of rivers.



Isotopic and elemental changes in seawater composition in the latest Quaternary: bottom, Ge/Si ratio ($\times 10^{12}$) of opal recovered from Antarctic piston cores²; middle, the SPEC-MAP time stack of oxygen isotope variations⁸; top, strontium isotope composition of foraminifera recovered from an equatorial Pacific piston core¹, with $\Delta^{87}\text{Sr}$ in parts per million equal to $\{[(^{87}\text{Sr}/^{86}\text{Sr})/0.709168]-1\} \times 10^6$.

Both records suffer from ambiguity. Detectable changes in the strontium isotopic composition of sea water might indeed be the result of adjustments in the total riverine flux of strontium to the oceans, because the ratio $^{87}\text{Sr}/^{86}\text{Sr}$ of average river water is significantly larger than that of sea water. On the other hand, they might be caused by changes in the relative contributions from weathering of areas of radiogenic (high $^{87}\text{Sr}/^{86}\text{Sr}$) and nonradiogenic (low $^{87}\text{Sr}/^{86}\text{Sr}$) rocks, changing the net composition of the riverine flux. These shifts would presumably require a geographic redistribution either of areas of high runoff, or of other factors that influence rates of

chemical denudation, such as land cover, temperature, or the exposure of fresh mineral surfaces (accelerated by the physical effects of glaciation).

The Ge/Si ratio of sea water is subject to the same factors (magnitude of riverine fluxes, composition of major source regions) as well as an additional one, the modification of Ge/Si ratios during weathering. Chemical weathering of silicate minerals is typically incongruent — that is, both solid phases (clays) and dissolved materials are produced. (During congruent weathering only dissolved materials are produced.)

The 2:1 clays (such as smectite) preferentially incorporate Ge, so that streams draining regions where smectite is forming tend to have smaller Ge/Si ratios⁴. (The additional complication of Ge/Si fractionation by marine diatoms^{5,6} can apparently be neglected according to diatom uptake experiments performed by Froelich *et al.*², although the appropriateness of applying their results, obtained from two species, to the exceedingly diverse natural system may be called into question.)

With the contradictions between the strontium-isotope and Ge/Si records, a simple explanation invoking either increases or decreases in the global river discharge or material flux is untenable. Instead, separate explanations have been advanced for the two oceanic records, linked only by their dependence on regional effects of global climate change. Froelich *et al.*² explain the trend towards lower Ge/Si ratios in glacial seas as a response to a shift in weathering regime, towards an increasing dominance of regions where incongruent weathering takes place (typically mountainous or glaciated regions with thin soils; the "weathering limited" regions of Stallard and Edmond⁷). Such a shift could result

1. Dia, A. N., Cohen, A. S., O'Nions, R. K. & Shackleton, N. J. *Nature* **356**, 786–788 (1992).
2. Froelich, P. N. *et al.* *Paleoceanography* (in the press).
3. Mortlock, R. A. *et al.* *Nature* **351**, 220–223 (1991).
4. Murnane, R. J. & Stallard, R. F. *Nature* **344**, 749–752 (1990).
5. Murnane, R. J. & Stallard, R. F. *Paleoceanography* **3**, 461–469 (1988).
6. Froelich, P. N., Mortlock, R. A. & Shemesh, A. *Global biogeochem. Cycles* **3**, 79–88 (1989).
7. Stallard, R. F. & Edmond, J. M. *J. geophys. Res.* **92**, 8293–8302 (1987).
8. Imbrie, J. *et al.* in *Milankovitch and Climate* (eds Berger, A. *et al.*) 269–305 (Reidel, Dordrecht, 1984).

either from a redistribution of regions of more intense runoff, so that these regions coincide with weathering-limited regimes, or from a change in weathering regime for those areas currently experiencing intense runoff. Dia *et al.*¹ suggest that an intensification of the Asian monsoon at the end of the last glaciation might have boosted the contribution of radiogenic strontium to the world ocean from the ⁸⁷Sr-rich Ganges/Brahmaputra river system.

Apparently, a unifying explanation for the link between climate and continental

weathering awaits a comprehensive, geographically based study. Despite the complexity of the current explanations, one message is clear: the ocean must be considered as an open system, receiving inputs from a variety of sources (dominated by rivers) and delivering material to a variety of sinks, on timescales as short as a few thousand years. □

Lee Kump is in the Earth System Science Center and Department of Geosciences, The Pennsylvania State University, University Park, Pennsylvania 16802, USA.

AGRICULTURE

Resisting resistance

Robert M. May

In the heady years after the Second World War it briefly seemed as if DDT and other pesticides would win the age-old human war on insect crop pests and disease vectors. Such dreams overlooked the way the natural world has been shaped by evolutionary challenges and responses — the challenge of chemical pesticides soon evoked the response of resistant genotypes of arthropods. Today, more than 500 species of insects and mites that are pests in crops or orchards have evolved resistance, often to a wide range of different pesticides¹. Overall, more than one-third of all agricultural production is still lost to pests, the same proportion as a century ago (although, of course, the total production is much larger than it was then)².

Now, a new battle-line is about to be drawn — genetically engineered crops that contain insecticidal toxins will soon be available commercially. But the usefulness of such transgenic crops will be short-lived if insects quickly adapt to the toxins. Hence the appearance of papers by McGaughey and Whalon³ and by Mallet and Porter⁴, who look ahead and ask how best to slow down the evolution of resistance.

Specifically, genes for *Bacillus thuringiensis* are being introduced into several crop plants. *Bacillus thuringiensis* itself is an aerobic, Gram-positive, spore-forming bacterium that is found commonly in the natural environment. It produces several toxins, notable among which are protein crystals (called δ -endotoxins) that are formed during sporulation. The genes encoding δ -endotoxin production have been cloned in other host bacteria, and thence transferred into crop plants. Compared with chemical pesticides, these toxins have the advantage that they are specific to pest insects. They do not persist in the environment, have no known adverse effects on mammals or birds, and often

will kill pest species without affecting the arthropod species that prey on them.

Bacillus thuringiensis has already been used directly, as an infectious disease agent or 'microbial insecticide', against several insect pests, notably the gypsy moth in North America. McGaughey and Whalon³ summarize eight studies of resistance to the bacterium's δ -endotoxins, and they give every reason to suppose that resistant genotypes will appear soon after transgenic crops incorporating these toxins are widely planted. Once the genes for resistance have appeared in the pest population, it will usually be only a few insect generations before the transgenic crops are being happily eaten by resistant pests.

More precisely, if we caricature resistance as arising at a single genetic locus with two alleles (R for resistant and S for susceptible), then we may call the population 'resistant' if the frequency of the resistance allele, R, exceeds 50 per cent. For insects feeding on toxic plants, the resistant genotypes (RR) may be taken to have relative fitness 1; the susceptible genotypes (SS) the diminished fitness $1-s$ (where s is the selection strength); and the heterozygotes (RS) some intermediate fitness $1-(1-h)s$ (where h is the degree of dominance of the resistance allele).

Starting from some very small initial frequency, ϵ (which may be one in a thousand or much less) for

the resistance allele, the time taken for the insect pest population to become resistant to the toxic plants is then

$$T(\text{res}) \sim T_g [\ln (1/2\epsilon) / \ln \{(1-(1-h)s)/(1-s)\}]$$

Here T_g represents the generation time of the insect pest (which may be a few weeks to a few months, or less commonly a year or more). Notice that $T(\text{res})$ depends directly on the generation time, T_g , but only weakly — logarithmically — on s , h and ϵ . This expression for $T(\text{res})$ is only an approximation. In particular, if R is completely recessive (so that $h=0$) then the denominator in the equation for $T(\text{res})$ tends to zero and we have $T(\text{res}) \rightarrow \infty$; even here, however, the equation is telling us something sensible, namely that $T(\text{res})$ is very long when R is perfectly recessive^{5,6}. These calculations are supported by evidence about the time taken (characteristically 5 to 20 generations⁵) for various pest populations to evolve resistance to chemical pesticides.

These remarks apply to the rate of evolution of resistance in more-or-less homogeneous situations. Things are more complicated if we have pests moving among regions where pesticides are selecting for resistance, and other regions which are untreated and where susceptible individuals flourish. Simple population-genetical models suggest that

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Bacillus thuringiensis — genes for the δ -endotoxins it produces are being engineered into crop plants.

the time taken for significant levels of resistance to become established in the treated region depends on the balance between the selective pressures exerted by the pesticides, and the amount of movement or 'gene flow' between treated and untreated regions: at high levels of gene flow, resistance takes a long time to become established; at low levels, resistance appears in a few generations; and at intermediate levels, there can even be, for quite a long time, two alternative states, one at high and the other at low levels of resistance (although this two-state regime is probably of more academic than practical interest)^{5,6}.

Building on these ideas, Comins⁶ and others⁷⁻⁹ have proposed that the time taken for resistance to appear can be significantly lengthened if fresh supplies of susceptible individuals keep appearing, each generation, in treated regions. These susceptible individuals could come from untreated refugia, deliberately set aside to conserve susceptibility, or they could (usually with more effort) be bred and released. McGaughey and Whalon refer to this as "resistance management". Other potential ways to the same end involve alternation or rotation of two or more insecticides or toxins (hoping that a significant degree of reversion to susceptibility will occur during the intervals when a particular agent is not in use), or simply using mixtures or 'stacks' of pesticides.

Intuition suggests that such methods of alternation or combination will typically result in resistance to the entire suite of insecticides or toxins evolving over much the same time-span as it would were the agents introduced seriatim. More detailed analysis of the population genetics broadly seems to support this intuition, although events can move faster or slower, depending on possible nonlinear couplings among resistance mechanisms and genes^{10,11}. Gene flow from refugia, on the other hand, can greatly lengthen the time before resistance becomes a problem.

What of other options? Might it be possible to create effective refugia for susceptible pest genotypes by using seed mixtures, with toxic and toxin-free plants in the same fields? Or even to retard the evolution of resistance by arranging for some parts of the plant tissue to be toxic and others not? Mallet and Porter⁴ use population-genetical models to analyse these options. Unfortunately, they find that, provided insects move from plant to plant — or, for tissue-specific toxins, from one part of a plant to another — "seed mixtures may actually hasten insect resistance compared with pure stands of toxic plants".

This perverse outcome is most likely to arise when resistance has low 'genetic

dominance' — that is, when the heterozygous genotypes, RS, have fitness not much greater than susceptible ones, SS (*h* is small) — which is exactly the circumstance under which gene flow from spatial refugia is likely to cause resistance to evolve very slowly. In contrast to the complicated effects of pest movement among plants grown from seed mixtures, untreated regions where susceptibility is conserved have the simple effect of reducing the overall average level of selection for resistance, without altering overall dominance.

Mallet and Porter conclude that seed mixtures or tissue-specific expression of toxins may well accelerate the evolution of resistance instead of retarding it, and that on present indications toxin-free refugia look like the best bet. They speculate that legislation, rather than exhortation for cooperation, may be the answer to the attendant problems of requiring cooperation among farmers. Such legislation could, for example, take the form of an expansion of the current EC 'set-aside' programme, under which land is left fallow to maintain crop prices.

Here I have dealt largely with the biological aspects of the evolution of resistance to insecticides and toxins. Ultimately, however, the discussion must be embedded within a larger economic and social setting — as in many other bioeconomic contexts¹², short-sighted strategies of application of chemical pesticides are often rooted in real differences in economic interests between pesticide manufacturers, and farmers and society more generally⁵. We can, we must hope, learn from that lesson as well as from the new work on the evolution of resistance to transgenic crops. □

Robert M. May is in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

1. Georgiou, G. P. in *Managing Resistance to Agrichemicals: From Fundamental Research to Practical Strategies* (eds Green, M. B. et al.) 18 (American Chemical Society, Washington, DC, 1990).
2. Pimentel, D. *Bull. ent. Soc. Am.* **22**, 20–26 (1976).
3. McGaughey, W. H. & Whalon, M. E. *Science* **258**, 1451–1455 (1992).
4. Mallet, J. & Porter, P. *Proc. R. Soc. Lond. B* **250**, 165–169 (1992).
5. May, R. M. & Dobson, A. P. in *Pesticide Resistance: Strategies and Tactics for Management* (eds Glass, E. H. et al.) 170–193 (National Academy Press, Washington, DC, 1986).
6. Comins, H. N. *J. theor. Biol.* **64**, 177–197 (1977).
7. Gould, F. *Bioscience* **38**, 26–33 (1988).
8. Roush, R. T. *Pesticide Sci.* **26**, 423–433 (1989).
9. Tabashnik, B. E. & Croft, B. A. *Envir. Entomol.* **11**, 1137–1148 (1982).
10. Via, S. in *Pesticide Resistance: Strategies and Tactics for Management* (ed. Glass, E. H. et al.) 222–235 (National Academy Press, Washington, DC, 1986).
11. Uyenoyama, M. K. *Pesticide Resistance: Strategies and Tactics for Management* (eds Glass, E. H. et al.) 207–221 (National Academy Press, Washington, DC, 1986).
12. Clarke, C. W. *Mathematical Bioeconomics* (Wiley, New York, 1976).

Tactosociology

LAST week Daedalus pointed out that each region of each person's skin carries a unique and stable ecology of harmless microorganisms. Any skin contact exchanges some of these organisms; the colonists persist for some time until ultimately ousted by the better-adapted natives. DREADCO's sociologists are now using these bacteriological fingerprints of touch to study human associations.

A group of people who regularly shake hands, for example, will come to hold a group of hand-inhabiting organisms in common. A bacterial culture from someone's hand could establish whether he belonged to the group. Social groups that go in for formal kissing will soon share a characteristic facial flora. Other skin organisms could reveal the social circles of closed or preferential sexual choice. Flora that can be transferred on currency notes should neatly trace commercial clusterings. Social groupings could thus be studied quite non-intrusively, by simple bacterial sampling.

Daedalus hopes to obtain even more information by introducing test organisms into society and following their spread. A well designed, more efficient skin ecology should permanently displace the previous flora. Once launched, it would propagate by contact as a harmless 'invisible epidemic', tracing out the complex networks of human touch. Quite unrecognized by its carriers, such an epidemic might show up in the pharmacy statistics. For we are all immune to our normal skin flora; they don't infect small abrasions. But a new skin flora would briefly outwit the immune system and infect such tiny cuts, provoking the purchase of sticking-plaster or antiseptics. A 'front' of such sales, looking like a travelling wave of accident-proneness, should accompany each epidemic.

By their rate of spread, the subgroups they infect and those they pass by, DREADCO's invisible epidemics will reveal the true divisions and groupings in society. For utopian politicians, they will provide the ultimate test of the classless society — in which an invisible epidemic would spread inclusively and smoothly through the whole population. Meanwhile, Daedalus has another political use for them. At the next election, he plans covertly to infect the rival teams of political activists with mutually incompatible skin ecologies. As they furiously kiss the babies and press the flesh of their constituents, each team will propagate its own invisible epidemic into its community of influence. DREADCO will then predict the election's outcome from a bacterial sampling of voters.

David Jones

Carbon onions in meteorites

SIR — Ugarte reported that soot, when annealed by intense electron radiation, naturally transforms itself into giant nested shells ('onions') of carbon¹. Kroto² in *News and Views* pointed out that these carbon onions could be related to fullerenes³ and 'nanotubes'⁴. The carbon onions are formed by irradiation-stimulated graphitization, which is different from the accretion mechanism originally considered in laser ablation³ and arc-discharge processes⁴ used for fullerene synthesis.

Some years ago, Smith and Buseck⁵ observed well-ordered graphitic particles in the Allende meteorite which were concentric circular or polygonal structures, about 100–500 Å in diameter. These particles were found in residue samples that had been etched in fuming HNO₃ within the matrix of the meteorite, which is predominantly amorphous carbon. These concentric structures observed in the meteorite are remarkably similar in size and appearance to carbon onions¹ and nanotubes⁴.

Ugarte¹ describes a transition of irradiated soot in a 300-kV high-resolution electron microscope, originally composed of nanometric needles and tubular particles, which slowly becomes more spherical with prolonged exposure to the beam. These spheroidal graphite shells are similar to the nanotube structures described by Iijima⁴, produced by an arc-discharge evaporation method similar to that used for fullerene synthesis. Detailed examination of the newly formed graphite onions created in the electron beam shows that they consist of an assembly of concentric graphitic cages, with the distance between layers similar to that for bulk graphite (3.4 Å). Electron micrographs of the nanotubes also correspond to the lattice images of graphite (3.4 Å). This is consistent with the observations of Smith and Buseck⁵ that the lattice fringe spacing of the particles identified in Allende is 3.4 Å, corresponding to partially graphitized carbon. In the light of these structural similarities, we propose that the structures identified by Smith and Buseck in Allende are related to the carbon onions and nanotubes, and perhaps fullerenes³. If this is the case, then the presence of these onion-like structures in Allende is the first observation of fullerene-related carbon clusters in an extraterrestrial sample.

A question that remains, however, is how could these carbon onions form in a meteorite such as Allende? We cannot exclude the possibility that the amorphous carbon in Allende was transformed during examination by the high-resolution microscope. But a natural

mechanism that could be responsible for the formation of concentric structures in carbon-rich meteorites is shock synthesis. Yamada *et al.*⁶ simulated shock conditions in the laboratory using acetylene polymer samples and examined the shocked products using the high-resolution electron microscope. Spherical particles, composed of graphitic layers with lattice fringe spacings and ring patterns corresponding to a graphitic reflection, were observed. In these experiments impact energies were of the order of 10¹² erg cm⁻³. Impacts delivering energies of this magnitude would exceed the mechanical strengths of planetesimals⁷ and lead to catastrophic fragmentation. Shock phenomena in the interstellar medium, particularly on grains containing organic molecules, could also drive synthesis of concen-

tric carbon structures⁸. Lower-energy impacts may be sufficient to drive synthesis of concentric carbon clusters, but this has yet to be demonstrated experimentally.

L. Becker

G. D. McDonald

J. L. Bada

*Scripps Institution of Oceanography,
University of California at San Diego,
La Jolla,
California 92093-0212, USA*

1. Ugarte, D. R. *Nature* **359**, 707–709 (1992).
2. Kroto, H. W. *Nature* **359**, 670–671 (1992).
3. Kroto, H. W., Heath, J. R., O'Brien, S. C., Curl, R. E. & Smalley, R. E. *Nature* **318**, 162–163 (1985).
4. Iijima, S. *Nature* **354**, 56–58 (1991).
5. Smith, P. & Buseck, P. R. *Science* **212**, 322–324 (1981).
6. Yamada, K., Kunishige, H. & Sawaoka, A. B. *Naturwissenschaften* **78**, 450–452 (1991).
7. Stoffer, D., Bischoff, A., Buchwald, V. & Rubin, A. E. in *Meteorites and the Early Solar System* (eds Kerridge, J. F. & Matthews, M. S.) 165–204 (Univ. Arizona Press, Tucson, 1988).
8. Kroto, H. W. in *Polycyclic Aromatic Hydrocarbons and Astrophysics* (eds Leger, A., d'Hendecourt, L. & Boccara, N.) 197–206 (Reidel, Dordrecht, 1987).

Forensic evidence by DNA sequencing

SIR — We wish to report on a rape case in which direct genomic sequencing was used to compare HIV-1 *pol* gene sequences from the male defendant and the female victim. We believe that this was the first time that evidence produced by DNA sequencing has been used in court.

The male was a Swedish intravenous drug abuser who was found to be HIV-1

to frozen peripheral blood mononuclear cells (PBMC) from the woman obtained 17 months earlier (5 weeks after the suspected transmission). Unrelated HIV-1 infected individuals (10 intravenous drug abusers and 20 homosexual men) from the same geographical area (Stockholm) served as controls. A region of the *pol* gene corresponding to amino acids 8–222 of the reverse transcriptase was

	160	170	180	190	200	210	220
Consensus	GWKGSFAIFQSSMTKILEPFRKQNPDIYIYQYMDLIVGSDLEIGQHRTKIEELRQHLRWGFTTPDKKHQ						
Male	-----4-----2-----					K-3-1-----	
Female	-----4-----2-----					5-K-----6-----	

Amino-acid sequences of the HIV-1 reverse transcriptase of the virus populations in uncultured PBMC from the male and the female aligned to a consensus of published HIV-1 sequences (Los Alamos database). Dash, identity with the consensus sequence; letter, amino-acid change; numbers, positions with genetic polymorphism within the virus population in the sample according to the code: 1=K, N, R and S, 2=I and M, 3=Q and E, 4=D and E, 5=I and V, 6=R and S. A nucleotide was scored as polymorphic only if at least 20% of the virus population was reproducibly estimated to be mutant.

positive in 1986. Before the present analysis was conducted, he had been convicted for rape and deliberate transmission of HIV-1 in the Stockholm district court. The evidence for this conviction did not include any forensic analysis because the woman involved did not file a report until a year after the alleged rape. However, she had been shown to seroconvert for HIV-1 antibodies within a few weeks of the rape. For this reason, a genetic analysis of the HIV-1 strains carried by both parties was performed before the case was tried in the court of appeal.

Two samples of whole blood were obtained on separate days from both subjects. In addition, we had access

directly sequenced from uncultured PBMC from the male, the female and the controls using a solid-phase sequencing method^{1–3}.

The direct DNA sequence analysis showed that the virus populations harboured by the male and the female were highly homologous (see figure). It was possible to identify a unique Arg 49, Arg 83, Asp 169, Met 178, Lys 204 and Lys/Ser 211 signature pattern which was shared between the male and female sequences. No other sequences from the intravenous drug abusers and homosexual men from the same geographical region fitted more than two out of these six amino acids. Phylogenetic tree analysis also showed that the sequences from

the male and the female clustered closely together. Similar patterns were observed with sequential samples from single individuals, but not with unrelated samples.

DNA sequencing was also used in the 'Florida dentist' case⁴, but the results were not used as evidence because the case was settled out of court. Still, it is interesting to compare the strategies used in these two investigations. One important difference is that we studied the *pol* gene instead of the more variable *env* gene. Because a relatively stable region was analysed, a rapid direct sequencing method could be used which eliminates the need to analyse multiple clones from each sample.

Based on the genetic analysis and other evidence in the case, the verdict from the district court was upheld in the court of appeal. Furthermore, the defen-

dant was not given leave to appeal to the supreme court. This study demonstrates the power of direct genomic sequencing in forensic medicine, but the method also has broad applications outside this field.

J. Albert

Department of Virology,

National Bacteriological Laboratory,

J. Wahlberg

M. Uhlén

Department of Biochemistry and

Biotechnology,

Royal Institute of Technology,

S-100 44 Stockholm,

Sweden

1. Hultman, T., Ståhl, S., Hornes, E. & Uhlén, M. *Nucleic Acids Res.* **17**, 4937–4946 (1989).

2. Wahlberg, J., Lundberg, J., Hultman, T. & Uhlén, M. *Proc. natn. Acad. Sci. U.S.A.* **87**, 6569–6573 (1990).

3. Wahlberg, J. *et al.* *FASEB J.* **6**, 2843–2847 (1992).

4. Ou, C.-Y. *et al.* *Science* **256**, 1165–1171 (1992).

lineages carrying sequentially arranged sister taxa (plesions) below the crown group⁷, whereas other groups seem to show a more symmetrical arrangement⁸. The patterns do not appear to be affected by the removal of recent taxa⁶, and the search for alternative topologies has not been strikingly successful^{4,5}. These are all interesting results, worthy of further study; Gee does them no service by conflating them with 'join-the-dots' transformation series.

Per Erik Ahlberg

Department of Zoology and University Museum,

University of Oxford,

South Parks Road,

Oxford OX1 3PS, UK

1. Gee, H. E. *Nature* **360**, 529 (1992).

2. Fox, R. C., Youzwyshyn, G. P. & Krause, D. W. *Nature* **358**, 233–235 (1992).

3. Panchen, A. L. *Zool. J. Linn. Soc.* **74**, 305–328 (1982).

4. Panchen, A. L. & Smithson, T. R. in *The Phylogeny and Classification of Tetrapods* (ed. Benton, M. J.) 1–32 (Clarendon, Oxford, 1988).

5. Godfrey, S. J. *Phil. Trans. R. Soc.* **B323**, 75–133 (1989).

6. Gauthier, J., Kluge, A. G. & Rowe, T. *Cladistics* **4**, 105–209 (1988).

7. Patterson, C. & Rosen, D. E. *Bull. Am. Mus. Nat. Hist.* **158**, 81–172 (1977).

8. Ahlberg, P. E. *Zool. J. Linn. Soc.* **103**, 241–287 (1991).

Therapsids and transformation series

SIR — In his discussion of the supposed Palaeocene therapsid fossil *Chronoperates*, Gee¹ presents a distorted picture of current phylogenetic practice among vertebrate systematists. He says that the "distinction between mammals and therapsids" is a side-effect of rearranging a bush of convergent mammal-like lineages, *post hoc*, into an ordered stem lineage. He also asserts that *Chronoperates* "is a fossil out of time only if one assumes that the transformation series between mammals and reptiles happened in the way the textbooks would have us believe".

This assertion is mistaken. There exists a clade Mammalia, characterized by the possession of certain derived features, whose earliest known representatives date from the Upper Triassic; *Chronoperates* (if correctly interpreted by Fox *et al.*²) lacks some of these derived characters and thus presumably represents a sister-group of the Mammalia; the last common ancestor of the mammals and *Chronoperates* must antedate the earliest known mammals, and is thus of Upper Triassic or earlier date; the occurrence of *Chronoperates* in the Palaeocene therefore implies the existence of an otherwise undocumented lineage extending through the Jurassic and Cretaceous. It is indeed a "fossil out of time".

The question of the "distinction between mammals and therapsids" is essentially a semantic rather than a phylogenetic problem, and has no real bearing on the phylogenetic or temporal position of *Chronoperates*. Gee accepts that the numerous derived mammalian characters are unlikely to have arisen simultaneously, and recognizes that this implies the existence of a stem lineage

showing a sequence of character acquisition. However, he seems to believe that the very existence of the crown group will prejudice the investigation, giving undue importance to those characters which would determine the position of the taxon on the stem. This idea has been pursued by Panchen³, who has explicitly tried to look for more 'symmetrical' alternatives to these orderly sequences of sister taxa. But the resultant cladograms⁴ are generally less parsimonious than their competitors⁵; it seems that the orderly appearance of derived characters is not an artefact.

Gee goes on to conflate this criticism with the unrelated and discredited practice of arranging fossil taxa in "ancestor-descendant" sequences. People who study the mammalian stem lineage have used cladistic techniques for more than a decade, and their phylogenetic hypotheses are falsifiable within that methodology. They are emphatically not looking for ancestors.

Gee has missed the point of the paper by Gauthier *et al.*⁶. It is not the mammals, but the 'therapsids' near the middle of the stem group which cause the flip from one tree topology to another. Removing the recent groups does not affect the outcome. The stem taxa change the topology because they demonstrate, through the combination of characters they possess, that the supposed mammalian-bird synapomorphies are in fact convergent. The methodology of Gauthier *et al.* is not unsound because it fails to give primacy to data from recent groups; rather, it demonstrates the untenability of this 'transformed cladistic' precept.

Cladistic analyses of several major vertebrate groups have revealed stem

SIR — The announcement of an anachronistic fossil — the mammal-like reptile *Chronoperates*¹ — and the ensuing arguments over its true identity^{2–4} do indeed open up old debates as Gee points out in News and Views⁵. The debates concern the difficulty of interpreting incomplete fossils, the nature of groups, and how we recognize their members.

As Rowe and Gauthier acknowledge⁶, no one has had difficulty in recognizing a Recent mammal (although there was a 'near miss' when the platypus was first brought back to Europe). Mammalian characters are obvious, and mammals can be diagnosed (distinguished) from their Recent sister-group (birds or birds+crocodyles). They are clearly monophyletic, with a single history which is separate from that of any other Recent group. Part of that history will be evidenced by animals now extinct and which, individually, possessed some, but not all, of the characters of modern mammals. These extinct animals are nevertheless mammals (see ref. 7) and can be arranged sequentially along an inferred stem lineage leading from the origin of mammals to the cladogenetic event which led to modern mammals (crown-group mammals). The fossils which may be arranged along this lineage are the traditional 'mammal-like reptiles'. Recent attempts to divide up this lineage arbitrarily into mammals and non-mammals on some 'key character' (see refs 6, 8) are subjective. *Chronoperates* is clearly a mammal as it shares a synapomorphy of mammals (a mas-

seter fossa developed on the dentary).

The difficulty of attempting to assign *Chronoperates* to a grouping of 'mammal-like reptiles' is that this latter is not a monophyletic group with a unique history. Instead, the 'mammal-like reptiles' are a paraphyletic assemblage. There is no character by which membership can be recognized. Mammal-like reptiles are mammals without the specializations of modern mammals.

The fact that *Chronoperates* lacks double-rooted cheek teeth, a cingulum or prismatic enamel tells us nothing since these are all primitive features, to be expected in a wide variety of early mammals as well as other amniotes. Fox *et al.*¹ further argue for the presence of postdentary bones based on the presence of a trough on the dentary. These, they say, would exclude *Chronoperates* from the mammals (the modern members of which have a single bone — the dentary — in the lower jaw) and justify its placement as a mammal-like reptile. There are two issues involved: the interpretation of the fossil specimen and the significance of the reconstruction. Fox *et al.*¹ believe that the fossil is incomplete and that the trough is evidence of other bones articulating at this point. But Sues³ argues that the fossil is intact, that the trough received soft tissue and that postdentary bones were not present. Absence in fossils is always difficult to explain because it can be either real or preservational. But even if Fox *et al.* were correct, this would not deny the mammalian affinities of *Chronoperates* because the existence of postdentary bones is the primitive mammalian condition.

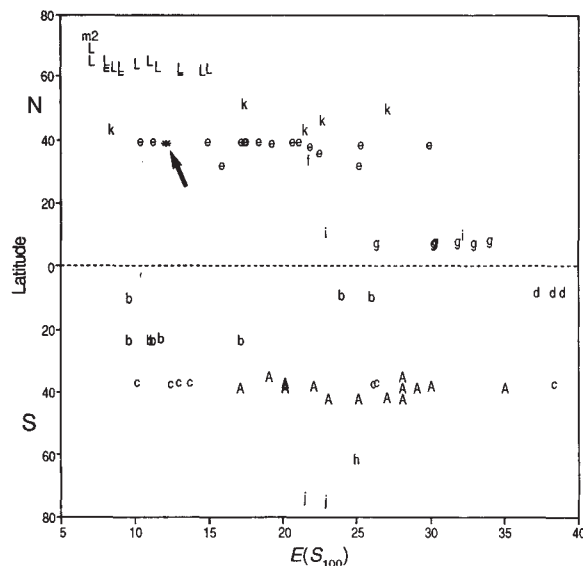
To argue that *Chronoperates* is a mammal-like reptile is a futile exercise as the claim can never be justified by observable characters, although it may be possible to show that it belongs to a monophyletic subgroup. Fox *et al.* are right to claim that *Chronoperates* belongs to the mammalian (their cynodont) clade, but exactly where it belongs depends on the recognition of synapomorphies with one or other of the included taxa.

Peter L. Forey

Department of Palaeontology,
The Natural History Museum,
Cromwell Road,
London SW7 5BD, UK

Marine species richness

SIR — Estimates for the number of species in the deep sea¹ have been discussed in News and Views². Grassle and Maciolek¹ reported on 14 samples taken within 176 km from the slope (depth range 1,500–2,500 m) of the north-western Atlantic where they found 798 species of invertebrates among 90,677 individuals in 21 m² of box-core samples. Using rarefaction methods and estimates of species change along spatial gradients



Expected number of species in 100 individuals ($E(S_{100})$) versus latitude. A, Southeastern Australia⁵; b, Angola Basin; c, Argentine Basin; d, Brazil Basin; e, Gay Head-Bermuda transect; f, Mediterranean Sea; g, Guiana Basin; h, Scotia Basin; i, Sierra Leone Basin; j, Weddell Sea; k, West European Basin⁴; L, Norwegian Basin; m, Greenland Basin³; asterisk, off New Jersey and Delaware (arrow, United States¹).

they calculated a global species deep-sea biodiversity of 10^7 invertebrate species. May² questioned these assumptions, specifically, that species replacement along global gradients is linear. He proposed that, because about half the species discovered were already known, the number is unlikely to exceed half a million.

We have independently collected data on isopods from the same depth range on continental slopes in several places in the Pacific and Atlantic oceans, and have information from the Arctic³ (see figure). These figures show that Grassle and Maciolek's data are far from typical and their species estimates are, in fact, at the lower end of those available from the oceans generally. Their estimate of the number of species of isopods expected to be represented in 100 individuals, $E(S_{100})$, is 12 whereas our data range from 7 in the Norwegian sea to 39 in the Argentine Basin. These data are variable

but show that the Norwegian and Greenland seas are relatively impoverished (as expected), that there are more species in mid-latitudes in both hemispheres, and that tropical slopes may be richer in species than temperate ones. The median $E(S_{100})$ of all data in the figure, Arctic species excluded, is 22 species.

These new data bring into question the appropriateness of using data from one site in the North Atlantic for global estimates. New calculations of isopod richness from samples taken during the Gay Head-Bermuda transect⁴ are consistently higher than Grassle and Maciolek's figure from the same region

but are on average lower than those for a similar latitudinal range off south-eastern Australia (manuscript submitted). The oceans of the Southern Hemisphere are much larger than those of the Northern Hemisphere and the North Atlantic is relatively young compared to the Pacific. Therefore, extrapolations to the global species number should include data from the South Pacific. We suspect new estimates could be much higher than so far admitted.

May's estimate from Grassle and Maciolek's data on the basis of fraction of known species may be unreliable. In Grassle and Maciolek's samples 31% of peracarid crustaceans were identified to species, a figure which might apply to isopods in the northwestern Atlantic but which is known to be

six times higher than in the Central Pacific. In southeastern Australia the equivalent figure is 10%. But isopods have received more recent taxonomic attention in Australia than most other invertebrate taxa and half of the known species have been described in the past 7 years by only three principal authors (G. C. B. P., N. L. Bruce and J. Just). The known fraction of other invertebrate species in this region is also consistently low, for example in aplousophoran molluscs⁵. May used a factor of 2 to extrapolate from known to total fauna; a factor exceeding 20 might be more reasonable for the oceans as a whole.

The discussion so far has been limited to data on the benthos of mid-slope communities, which are only a small part of the ocean floor. Equivalent data from the oceanic basins are scarce but do suggest that species richness measured in

1. Fox, R. C., Youzwyshyn, G. P. & Krause, D. W. *Nature* **358**, 233–235 (1992).
2. Novacek, M. J. *Nature* **358**, 192 (1992).
3. Sues, H.-D. *Nature* **359**, 278 (1992).
4. Fox, R. C., Youzwyshyn, G. P. & Krause, D. W. *Nature* **360**, 540 (1992).
5. Gee, H. *Nature* **360**, 529 (1992).
6. Rowe, T. & Gauthier, J. *Syst. Biol.* **41**, 372–378 (1992).
7. Ax, P. *The Phylogenetic System: The Systematization of Organisms on the Basis of their Phylogenesis* (Wiley, New York, 1987).
8. Lucas, S. *Syst. Biol.* **41**, 370–371 (1992).

the way outlined here may be higher than on the slope. For example, $E(S_{100})$ values of up to 48.9 isopod species have been recorded from the central Pacific Ocean⁶.

Grassle and Maciolek also suggested that shallow marine environments outside of the tropics have relatively few benthic species compared to the deep sea. They cite 200 species as the number expected for large collections from the Georges Bank. In southeastern Australia the fauna is far richer: more than 700 infaunal benthic species in Port Phillip Bay⁷; more than 800 species in 10 m² of benthos in Bass Strait (unpublished data). This is further evidence that the North Atlantic is not typical of oceanic biodiversity and that further exploration in the little-studied Southern Hemisphere will significantly add to estimates of marine species richness.

Gary C. B. Poore

Department of Crustacea,
Museum of Victoria,
Swanston Street, Melbourne,
Victoria 3000, Australia

George D. F. Wilson

Division of Invertebrate Zoology,
Australian Museum,
PO Box A285, Sydney,
New South Wales 2001,
Australia

MAY REPLIES — Poore and Wilson make an important contribution to discussions of benthic macrofaunal diversity. They emphasize that benthic fauna can vary greatly from place to place, so that estimates of total species numbers based on patterns found in any one location can easily be confounded by different patterns in other locations. Poore and Wilson suggest an overall factor of 20 might be appropriate in extrapolating from known to total benthic faunas,

leading to an estimate of around 5 million species. This is ten times my rather conservative guess², but still well below that in Grassle and Maciolek's important and provocative paper¹ (they estimated a few hundreds of millions, but scaled it back to 10 million or so to express their view that species numbers are likely to be lower on the floors of ocean basins than on continental shelves or mid-slopes; note that Poore and Wilson believe, to the contrary, that species richness may be higher in ocean basins).

Poore and Wilson use $E(S_{100})$, the average number of different species found in a sample of 100 individuals, as a guide to species numbers, inferring that more species are present when $E(S_{100})$ is higher. But $E(S_{100})$ measures evenness, not species richness. Very often a larger value of this index will indicate more species are present. But it is perfectly possible for two communities to have very different values of $E(S_{100})$, even though both have the same total of species (just as two countries with similar total populations and total wealth can have the wealth distributed in very different ways, from egalitarian to feudal extremes).

Robert M. May

Department of Zoology,
University of Oxford,
South Parks Road,
Oxford OX1 3PS, UK

1. Grassle, J. F. & Maciolek, N. J. *Am. Nat.* **139**, 313–341 (1992).
2. May, R. *Nature* **357**, 278–279 (1992).
3. Svavarsson, J., Brattegard, T. & Strömberg, J.-O. *Prog. Oceanogr.* **24**, 297–310 (1990).
4. Hessler, R. R. & Wilson, G. D. F. in *Evolution, Time and Space: The Emergence of the Biosphere* (eds Sims, R. W., Price, J. H. & Whalley, P. E. S.) 227–254 (Academic, London, 1983).
5. Scheltema, A. H. *Bull. Mar. Sci.* **47**, 50–61 (1990).
6. Thistle, D. & Wilson, G. D. F. *Deep-Sea Res.* **34**, 73–87 (1987).
7. Poore, G. C. B., Rainer, S. F., Spies, R. B. & Ward, E. *Fish. Wildl. Pap., Vict.* **7**, 1–78 (1975).

result would imply that it is highly unlikely that the global mean temperature depends on this cycle length.

Neill S. Cooper

104 Westhall Road,
Warlingham,
Surrey CR6 9HD, UK

1. Friis-Christensen, E. & Lassen, K. *Science* **254**, 698–700 (1991).
2. Laciš, A. A. & Carlson, B. E. *Nature* **360**, 297 (1992).
3. Kelly, P. M. & Wigley, T. M. L. *Nature* **360**, 328–330 (1992).
4. Schlesinger, M. E. & Ramankutty, N. *Nature* **360**, 330–333 (1992).
5. Cooper, N. S. *Nature* **296**, 131–132 (1982).
6. Chapman, S. & Lindzen, R. S. *Atmospheric Tides* (Reidel, Dordrecht, 1970).
7. Lean, J. *Science* **244**, 197–200 (1990).

Purity of nature?

SIR — The description by G. H. Gonnet and S. A. Benner (*Nature* **361**, 121; 1993) of their search of the one-letter coded sequences of proteins for long English words stirred memories of 32 years ago when I was writing a review on peptides and proteins for the Annual Reports of the Chemical Society. The one-letter code had recently been introduced by the late Professor F. Šorm of the Czech Academy of Sciences. Not many protein sequences were then known but they were clearly going to appear at an ever-increasing rate.

I enquired of the editor if I might use Šorm's code to save space; permission was withheld on the grounds that obscene sequences might be forthcoming. I wrote to Professor Šorm telling him of the decision and my frustration. He replied that he had written out all the sequences known at that time and had not found any sequence that could be construed as obscene. With tongue in cheek, he attributed this to the purity of nature.

In fact, with a characteristic versatility, he displayed a good working knowledge of the English vernacular and avoided the use of letters such as u, o and b, which might have given rise to obscene sequences. There are, however, some indelicate words that can be constructed from the partial alphabet used by Šorm.

With the immense databank of protein sequences now available, it would be interesting to know if Šorm's attribution of verbal purity to nature is really true. Whatever the answer, it appears from the two recondite examples cited by Gonnet and Benner that nature indulges in sesquipedalianism. Perhaps this is a reflection of the tendency of naturally occurring macromolecules to contain redundant sequences.

D. T. Elmore

Dyson Perrins Laboratory,
South Parks Road,
Oxford OX1 3QY, UK

Solar output and global warming

SIR — The debate as to whether solar variability is a significant cause of the rise in global mean surface air temperature^{1–4} has been based on data analysis. Although theoretical mechanisms have been postulated, they are difficult to test because of the lack of historical data on the variability of the solar output.

However, there is a terrestrial parameter which is highly sensitive to variations in the solar ultraviolet radiation — the atmospheric semi-diurnal tide⁵. Unlike the ocean tides, the atmospheric tides are predominantly forced by solar heating⁶. In particular, the semi-diurnal (12-hour) tide is driven by absorption of ultraviolet radiation in the stratosphere and thus is sensitive to wavelengths where solar output varies significantly⁷.

Historical data on the amplitude of the atmospheric semi-diurnal tide could therefore provide an indication of the magnitude of solar variability with sunspot cycle length.

Such a dataset is available as the semi-diurnal tide dominates the variations in atmospheric pressure in the tropics⁶. Hourly pressure data (required to calculate the tide) over many decades are available for various tropical meteorological stations.

I would like to suggest that someone who can access these data calculates the variation in the annual mean semi-diurnal tide in the tropics over the past 100 years. According to the above theory, this should show whether or not the solar ultraviolet output varies with the sunspot cycle length. A negative

Mind over body

Anthony Storr

A History of Hypnotism. By Alan Gauld. Cambridge University Press: 1992. Pp. 738. £75, \$140.

TWENTIETH-century scientists have, with some notable exceptions, fought shy of hypnosis and its associated phenomena. Like extrasensory perception, hypnosis has become linked with charlatans, conjurers, quacks, mediums and theatrical performances. As a method of treatment for neurotic disorders, it is recurrently revived. Some have found it useful as a technique for recovering buried memories in post-traumatic stress disorders of the kind that follow battles, air crashes and other disasters. But since 1892, when Freud gradually began to abandon hypnosis, mainstream psychiatry has generally eschewed it. Certain phenomena associated with hypnosis remain unexplained, however.

Dr Gauld, a psychologist at the University of Nottingham, England, has produced a work of reference on hypnosis that will be found to be indispensable by medical historians, by philosophers concerned with the history of ideas and by psychiatrists who are prepared to learn from the past, question their current assumptions and appreciate that the theories of today originate from the superstitions of yesterday. This detailed historical survey, replete with notes and references, ranges from the appearance of Mesmer in the late eighteenth century to the present day. As a chronicle of how our forebears thought about mind and body, Gauld's account is excellent. His remarks on hypnotic phenomena in the light of modern knowledge are less satisfactory.

Hypnosis purports to produce a state of mind seemingly intermediate between sleeping and waking in which the subject becomes particularly suggestible. In this state, subjects may recall memories to which they appear to have no access when normally conscious. They may also become insensitive to pain. Hypnotized subjects may be able to perform feats of physical endurance that they could not normally achieve. Some types of illness, especially those labelled 'psychosomatic', can be alleviated by hypnosis. The skin is an organ in which visible changes can be produced by hypnosis; a few varieties of skin disease respond favourably. Some subjects can be persuaded that they have regressed to childhood, and behave accordingly. Others seem to have hallucinations of objects or fail to perceive real objects when told that there is nothing to see.

Most subjects when they return to

normal consciousness are unable to recall anything that happened during the hypnotic period. Actions carried out as the result of post-hypnotic suggestion are either rationalized or unexplained. Study of the hypnotic state is relevant to sleepwalking and to cases of so-called 'multiple personality'. It is also relevant to some trance mediums who claim to be possessed by a variety of spirits. For once, Gauld has omitted an important

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Animal magnetism — a doctor uses Mesmer's technique by passing his hands over his patient.

study. He is unaware that C. G. Jung's dissertation for his medical degree was a detailed study of a medium, in which the origin of the spirits who supposedly possessed her is interpreted in terms of multiple personality and her own psychopathology.

Gauld plausibly suggests that the hypnotic state may itself be illusory, the product of historical misconception, compliance and the subject's expectation. Suggestion works whether or not a subject is hypnotized. Gauld does not refer to experiments in sensory deprivation that demonstrate that restriction of input — sound-proofing, darkness and so on — enhances suggestibility because the deprived subject pays especial attention to any stimulus that is allowed to reach him. This is surely relevant to hypnotic procedures that diminish the subject's awareness of any stimuli other than those provided by the hypnotist.

Similar considerations apply to hypnotic analgesia. As Gauld points out, the accounts of Esdaile's operations on hypnotized patients in India in the 1840s are so numerous and so well-authenticated that fraud can be dismissed. But he does not relate this form of analgesia to the well-known cases of fighting soldiers who have received dreadful wounds without realizing that they have been injured. Nor does he mention acupuncture. In both instances, analgesia is related to the direction of the subject's attention. This in no way conflicts with the notion that hypnosis may facilitate the production of the patient's own internal chemical analgesics, the endorphins.

When discussing post-hypnotic suggestion, Gauld appears to have difficulty in accepting the idea of unconscious estimation of time. But many people can wake themselves early by specifying a time before going to sleep. Nor does he relate the hypnotic state to one that he must have experienced: the normal phenomenon of being bored in a lecture and discovering that a whole host of mental images proceed involuntarily that are difficult to recapture when one's attention is once more engaged. We are all 'state-dependent' amnesics for much of what continuously goes on in our minds. In addition, we all have latent potentials that either suggestion or example may make actual. The hypnotized subject who performs physical feats that he could not carry out consciously is matched by the athlete who discovered that he could run a mile in four minutes only when Roger Bannister's example acted as a positive suggestion.

There is no comparison of hypnotically induced 'negative hallucination' (not seeing objects that are there) with the phenomenon of 'blindsight', in which brain-damaged individuals can be proved to be aware of objects that they claim they cannot see.

Gauld's most interesting idea is that the positive effects of hypnosis and other forms of psychotherapy may be mediated by the immune system. The electroencephalogram shows no difference between the brain waves of those who are deeply hypnotized and those who are awake. But perhaps the levels of hormones, neurotransmitters and other chemical substances, including the endorphins, may differ in the two groups. We shall have to wait and see. □

Anthony Storr, emeritus fellow of Green College, Oxford, is at 45 Chalfont Road, Oxford OX2 6TJ, UK.

Gravity defied?

Marie Boas Hall

The Janus Faces of Genius: The Role of Alchemy in Newton's Thought. By B. J. T. Dobbs. Cambridge University Press: 1992. Pp. 359. £30, \$47.95.

FEW men have had the evolution of their ideas examined so closely and intensively as Isaac Newton. Initially, historians concentrated on the origins of his singular achievements in natural philosophy, gradually basing their studies more and more on the enormous mass of manuscripts devoted to it. More recently, they have dealt with his nonscientific interests — alchemy, Biblical studies, theology — becoming less interested in how he formulated his conclusions than why.

For this, his manuscripts provide an exciting although puzzling source whose range and bulk is extraordinary — Newton was a man who always read and thought with pen in hand, writing and rewriting, taking notes and copying extracts even from books he owned, applying his profound mathematical mind to unravelling a host of obscure texts. Like very many of his contemporaries he believed firmly in the wisdom of the ancients, a wisdom to be understood by patient study. Mysterious, esoteric writings held a great fascination for him, whether the obscurities of the Book of Revelation, the problems of Biblical exegesis, or the deliberately unintelligible language of Hermeticism and alchemy. Possessed of a firmly rational mind, he could never believe that other men would have wasted their lives on chimeras; if they claimed to have discovered universal truth, they must have done so, and he was prepared to struggle all his life to understand the truth that must lie behind the puzzling symbolic language.

But to what end? Dr Dobbs, following the conclusion she tentatively reached in an earlier book in 1975, believes that he sought in alchemy an answer to the major remaining puzzle of his system of natural philosophy, namely the cause of gravity, accepting the alchemists' active vegetative spirit (supposedly responsible for change in both chemical and living matter) as the "hitherto unknown force" at which he hinted in the preface to the *Principia*, a force that might govern "all the phenomena of Nature". Historians have long explored the changing views held by Newton on the cause of gravity, ranging from unknown forces of attraction and repulsion to aethers of varying characteristics. Some scholars have accepted the alchemical interpretation. Many more have concluded that Newton really believed that the ultimate cause of



UNNATURAL history — some mediaeval zoologists were evidently defeated by the elephant, although this much caricatured creature had been known since Roman times. In this pair of pictures from about 1230, showing one in use in war and a group attending a fallen friend, the unfortunate beasts seem rather to resemble pigs. The invented scenes are reproduced in *Images of Science: A History of Scientific Illustration* by Brian J. Ford. Published by Oxford University Press (US) and The British Library (UK), \$45, £25.

gravity was the immanent will of God, as he said at the very end of the *Principia*. Now Dobbs agrees with this view. Although the first half of *The Janus Faces of Genius* is concerned with Newton's study of alchemy, the second and longer part deals with his varied studies in ancient philosophy (Pythagorean, Epicurean, neo-Platonist, Hermetic and Biblical) and his theological studies (especially his conversion to Arianism). The relevance of his Arianism apparently lies in the Arian belief that Christ, while divine, is the mediator between God and the material Universe, including man; for God, being non-material, does not act directly on matter (a view that undoubtedly subsumes other heresies besides Arianism).

In the modern way, Dobbs explores minutely the various strands of thought which, she finds, led Newton to his conclusions. She rightly points to the fact that Newton always sought rational not mystical truth, albeit often transcendental truth. She will not be surprised that I find the second half of her study more convincing than the first; I do not accept her belief that Newton was a "practising alchemist", not least because this belief rests partly on dubiously Newtonian texts that are most probably summaries

of the work of others. That Newton believed he understood the alchemical texts he studied so assiduously is obvious, if only from his use of alchemical names for (real) chemical substances in the chemical notebooks, so clearly experimental. Nor does it seem to me surprising that Newton should have endeavoured over so many years to solve such intellectual puzzles as the meaning of alchemical, Hermetic and Biblical texts: he was after all the most brilliant mathematician of his age and most mathematicians are inveterate puzzle solvers, who need to occupy their puzzle-solving facility when not involved in mathematical invention. Whether Newton would have agreed with Dobbs's analysis of how and why he arrived at his conclusions about universal gravitation and other physical forces is a question beyond solution. This is a complex and detailed work of scholarship, whose author has left no clue unexamined. Whether her proof is satisfactory must be left to the reader. □

Marie Boas Hall is at 14 Ball Lane, Tackley, Oxfordshire OX5 3AG, UK, and is the author of the recently published *The Library and Archives of the Royal Society, 1660–1990* (Royal Society).

NATURE · VOL 361 · 18 FEBRUARY 1993

Engineering revolution

Willem Hackmann

The Third Man: The Life and Times of William Murdoch 1754–1839, the Inventor of Gas Lighting. By John Griffiths. Andre Deutsch: 1992. Pp. 373. £20.

THIS book is not set in post-war occupied Vienna, but in the Cornwall and Midlands of the Industrial Revolution. William Murdoch is the Harry Lime of this plot, one of the Industrial Revolution's most creative figures. His co-stars are Matthew Boulton and James Watt, the partners of the innovative Soho Manufactory outside Birmingham where Watt's steam engines, and a host of other metal goods, were produced by the latest technical methods, many of them developed by Murdoch.

In *The Third Man*, Holly Martins, a second-rate writer of cowboy stories, tries to discover what had happened to Harry Lime. In this book, John Griffiths attempts to piece together the life of Murdoch from the scraps that have survived: a handful of letters and mechanical drawings, and a few objects. A sinister reason suggested by the author as to why so little of the correspondence appears to have survived of this hero of the Industrial Revolution, is that the records were 'cleaned' by Watt's relatives, not long after Murdoch's death in 1839. The author's finger of suspicion is pointed at James Watt junior, whose relationship with his father's erstwhile employee and friend was complex, and who was keen to sustain the 'Wattolatry' that had sprung up and which still dominates history books. In an attempt to correct this image, an exhaustive search was made of existing archives for Murdoch material, apart from in the Watts papers at Doldowld to which the author was refused access, although they are listed in the Public Record Office. This is unfortunate, even if the present owner, Lord Gibson Watt, took the trouble to assure the author that these papers contained no relevant material.

Watt deserves his reputation as one of the giants of the Industrial Revolution, which was built on his improved steam engine and on the flow-line production techniques pioneered by Boulton. In many respects Watt was a modest man but, by means of his voluminous correspondence, the author shows that he could be inordinately jealous of his reputation as an inventor. However, this was not an uncommon concern in a period noted for financial freebooting and heated patent litigations. The firm of

Boulton and Watt was continually trying to curb its competitors, either in court or out of it, and was not adverse to a little industrial espionage. The author recounts several humorous anecdotes of Murdoch's forays into enemy territory.

On one occasion he and Abraham Storey, manager of the foundry of the Soho Works, were sent to the famous Round Foundry in Leeds of Matthew Murray (who had become Boulton and Watt's principal competitor in the manufacture of steam engines) to discover how it could make the engines at a much reduced price. On this ostensible 'goodwill mission' they were admitted to every part of Murray's factory and succeeded in extracting from him the secret of his forge work in the course of two evenings of heavy drinking. Next, one of Boulton and Watt's former employees, who had been enticed by Murray to Leeds, tipped off the Soho firm about the improved casting techniques used by Murray in the manufacturing of his engines. In the meantime Murray had become furious with Boulton and Watt, for his expected return visit never materialized, no doubt because they feared that he might return the favour and spy on them. Murdoch was again sent to Leeds, but this time to make contact with their former employee who, after a drinking bout, agreed to accept a guinea a week to act as a spy until Watt junior decided that he should re-defect to the Soho firm. Through these tactics Murdoch discovered the casting techniques employed by Murray.

Murdoch emerges from these pages as a sympathetic figure, large in body and spirit, growing increasingly irascible with age, but who maintained his inventive mind virtually until the end. This Scottish miller's son, born and bred on the estate of James Boswell's father, had such an urge to make his way in the new technological age that, when a young man, he walked the 250 miles to Boulton's Soho Works. He remained associated with the firm until his seventies, for many years as installer and maintainer of its steam engines sold to the Cornish mines, and then as the manager of the foundry of the Soho Works. He made numerous inventions: the 'sun and planet gear' for Watt's steam engine, which turned its oscillation into rotary motion; the oscillating steam engine (later used on early steam boats); the D-slide steam valve; a model steam carriage; the sending of messages by compressed air; and, the cause of a great economic revolution in the nineteenth century, the harnessing of coal gas for the purpose of lighting.

Boulton died in 1809 and Watt ten years later. Murdoch was an inconspicuous spectator at the public meeting held in 1824 to raise a memorial sub-

scription to Watt for his bust to be placed in Westminster Abbey. On Murdoch's death no memorial was raised and no notices were published in national newspapers. A marble bust was made by Chantry (who also made busts of Watt and of that other important figure of the Industrial Revolution, the civil engineer John Rennie). Murdoch lapsed rapidly into an undeserved obscurity, overshadowed by the men for whose success he had been essential. This is well brought out in this book. Memorable in *The Third Man* was Anton Karas's evocative zither music. Memorable in Griffiths's biography is his description of the harsh working conditions of the Cornish miners among whom Murdoch worked for so many years. □

Willem Hackmann is at the Museum of the History of Science, University of Oxford, Old Ashmolean Building, Broad Street, Oxford OX1 3AZ, UK.

Clues to early life

Ian J. Fairchild

The Proterozoic Biosphere: A Multidisciplinary Study. Edited by J. William Schopf and Cornelis Klein. Cambridge University Press: 1992. Pp. 1,348. £105, \$195.

Two billion years, from roughly the fiftieth to the nineteenth percentiles of Earth history, is the compass of the Proterozoic Eon. At its dawn was the great protracted period of formation of the continental crust. Its demise was hastened, so we believe, by the accumulation of oxygen in the atmosphere to a degree that allowed all the basic body plans of complex life forms to spring forth in the early Cambrian burst of evolution.

The scientific study of the Proterozoic biosphere was slow to develop. Lurking at the base of the stratigraphic column was a mass of deformed and molten rocks. Far from revealing a coherent story of the early evolution of life, it was difficult to imagine fossils preserved at all. Yet at the turn of the century, Charles Walcott's acute eye had spotted the proverbial needle in the haystack in the form of carbonaceous spiral impressions in the kilometres-thick Belt strata of Montana, now known to be around 1,400 million years old. Such early discoveries remained in limbo for want of adequate biological or temporal context. S. A. Tyler and E. S. Barghoorn's description of silicified microbes from the early Proterozoic Gunflint chert in 1954

could have been achieved using the techniques available in Walcott's time, but a decade more passed before more detailed studies led to acceptance of these fossils' age and biotic origin.

Bill Schopf was in at the start of what he calls "the emergence phase" of this new science of Precambrian micropalaeontology in 1965. Fifteen years ago, he organized the Precambrian Paleobiology Research Group (PPRG), the endeavours of which led to a monographic publication on Archaean and early Proterozoic palaeobiology in 1983 and which, now enlarged to 41 experts, has produced this 3.4-kg epic. It contains printed summaries of several (bio)geochemical and palaeontological databases from both published work and results from 1,800 new samples from sites worldwide. Thence arise numerous synthetic articles on most aspects of the Proterozoic biosphere with an update on the Archaean, and a hard look at the Early Cambrian. Palaeontology is set in the context of changing atmospheric and oceanic chemistry, evolving styles of preservation of sediments in basins, and reconstructed continental positions. Considerable attention is given to proxy indicators as pointers to Earth's evolution, such as taxa-specific organic molecules extracted from Proterozoic sediments, stable isotope records and sediment geochemistry, and to molecular phylogeny of extant organisms as indicators of their relative times of emergence in the past. The multidisciplinary approach is exemplified by the devotion of nearly 100 pages to studies of modern microbial mat communities, relevant insofar as they model the Proterozoic benthos.

This book is not the place for wholly new theories, but contains a painstaking testing and development of numerous hypotheses, with a willingness to discard old favourites. Thus G. Vidal and A. H. Knoll's theory of late Proterozoic decline in the diversity of eukaryotic plankton is confirmed in modified form. Conversely, cyanobacteria now seem *not* to have evolved (we can point to a third of Proterozoic prokaryotic taxa still living today): cue Schopf's eulogy on their indestructibility and immutability. The immense co-operative effort represented by this work makes for a coherent text, satisfying on one level, although perhaps a bit too comfortable on another. Only rarely do subtle differences of emphasis emerge. Was there measurable oxygen in the Archaean atmosphere? Did any of the late Proterozoic Ediacaran animals survive into the Cambrian?

Despite its encyclopaedic scope, the tome is cautious in outlook, for several reasons. One can cite the morphological simplicity of most Precambrian fossils, problems in determining which struc-

tures actually are fossils, and the decreasing preservation potential with age of unmetamorphosed sediments. The imprecision of age determinations is a major stumbling block, and no radiometric data are gathered in this work. Other gaps are in sedimentary facies and petrology, stromatolites, and chemo- and lithostratigraphy (note the scope for correlation afforded by the low-latitude glacial deposits); but these are covered by other international groups whose work is in progress. And as the book is largely three years out of date, one must ask why all those tables weren't

printed on microfiche. The future of the full databases and their accessibility is also unclear. But let's not quibble: the *Proterozoic Biosphere* is undoubtedly a benchmark, as was intended. □

Ian J. Fairchild is in the School of Earth Sciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK.

■ Of related interest is the newly published *Fossil Prokaryotes and Protists* edited by Jeve Lipps. The book aims to provide a "consistent and comprehensive" text for courses in micropalaeontology. Blackwell, £35 (pbk).

Remembrance of things past

Paul G. Bahn

Archaeology: The Science of Once and Future Things. By Brian Hayden. W. H. Freeman: 1992. Pp. 484. \$23.95, £18.95 (pbk).

IN the esoteric and competitive world of introductory archaeology textbooks there are three basic models: the thorough survey of methods and techniques, the global chronological gallop through past cultures and the volume that aims to do both. As a coauthor, with Colin Renfrew, of an example of the first model (*Archaeology: Theories, Methods and Practice*, Thames and Hudson, 1991 — see *Nature* 354, 198; 1991), I suspected that this review would have to begin with a declaration of interest. But this is a different kind of book altogether.

Although resembling other introductory texts in size and format, Brian Hayden's book is actually a long, personal statement of what archaeology is about, and what it can tell us about the human condition in the past, present and future. His stated aim is to give students what they want to know, to present them with an archaeological perspective on the world that will last beyond exams.

He begins by setting out a few basic concepts, some of which will certainly challenge the minds of beginning students: that archaeology has no established facts and can never be truly objective; the purpose of typology; and the simple theories that underpin archaeological interpretations and explanations of cultural change. In the second section, the nucleus of the book, he leads us through human cultural development, from hunter-gatherers to the rise of food production, complex societies and civilizations. In the brief third section, he takes the story through the industrial and nuclear ages into the future, and philosophizes about our destiny.

Hayden's aim of establishing an "appropriate framework of conceptual

tools" and giving a "panoramic view of past events" is accomplished admirably, but at some cost. In contrast to textbooks that try to set out the data together with a variety of possible interpretations, he has been forced to make only very occasional reference to hard archaeological evidence, and to devote most of the book to telling a story — albeit a sometimes convincing one — at the most general level. His chosen explanations of change are determinedly materialistic, highlighting competition, rivalry, violence, warfare and Big Men. Perhaps a career in academic archaeology exposes one to these phenomena among one's peers to such a degree that they may come to dominate one's view of the past as well as the present, overshadowing any possible evidence of cooperation and spirituality.

Hayden's examples are often drawn from the areas he knows best: Mesoamerica, Australia and America's northwest coast. His bibliography is extraordinarily eclectic, from a history of square dancing to Tom Robbins's *Jitterbug Perfume*. The data provided are sound on the whole, although there are occasional lapses: for example, using an Ice Age engraving of two women as evidence for copulation, attributing the burial of Pompeii to an eruption by Mount Etna, and apparently fusing Gerald Hawkins with Stephen Hawking. Moreover, it is puzzling that a text that seeks to spare students unnecessary jargon and technical detail should subject them repeatedly to obscure terms such as chthonic or scrying.

Nevertheless, this highly readable and stimulating text constitutes an invaluable source of topics for discussion and will make its readers think. There can be no higher praise than that. □

Paul G. Bahn is a freelance writer on archaeology at 428 Anlaby Road, Hull HU3 6QP, UK.

The powers and pitfalls of parsimony

Caro-Beth Stewart

Parsimony analysis is a powerful tool for the study of biological evolution. It is used to construct phylogenetic trees, to evaluate alternative hypotheses objectively, and to study evolutionary pattern and process. Yet, as comparative data sets expand, the pitfalls of parsimony analysis are catching experts and novices alike.

THE principle of parsimony states that the simplest explanation consistent with a data set should be chosen over more complex explanations, and is a guiding tenet in scientific study (for critical review, see ref. 1). Parsimonious reasoning is a fundamental way of 'knowing' in comparative evolutionary biology, whether the raw data are molecular, physiological, morphological or behavioural¹⁻³. In particular, the principle of parsimony is the foundation of a powerful method, called parsimony analysis, that allows reconstruction of the past to be undertaken with logical and statistical rigour⁴⁻⁶. Here I outline the powers and pitfalls of parsimony analysis, with emphasis on its application to molecular sequence data⁷⁻¹¹. The following represents my opinion as regards a reasonable consensus in the field; it is written with the novice in mind.

Darwin¹² referred to evolution as 'descent with modification', and this simple phrase still embodies our best current understanding of the process^{1,4,13}. The goal of building phylogenetic trees is to discover the genealogical relationships between 'taxa' (biological entities such as genes, proteins, individuals, populations, species, or higher taxonomic units). A phylogeny is a hierarchy of nested sets of taxa indicating relative recentness of common ancestry^{4-6,12-15}.

A widespread misconception, especially concerning molecular sequences¹⁶, is that the building of evolutionary trees simply requires the grouping of taxa according to overall similarity^{1,4,13}. Several methods that embody this concept have been invented: they are referred to as distance-matrix methods^{9,17-20}. These methods ignore the possibility that apparent overall similarity and true evolutionary relationship are not necessarily the same thing^{1,4-6,21}. Two taxa can appear quite similar to each other yet be related relatively distantly. (After all, who among us would assume that an Elvis Presley look-alike is actually his twin brother?) Conversely, two closely related taxa may appear quite different from each other.

This distinction can be illustrated by the genealogical relationship of tetrapods (the four-legged vertebrates) to their fish-like ancestors. In Fig. 1, we consider the possible evolutionary relationships between three vertebrate lineages, one leading to sharks, another to lungfishes, a third to primates. If these species were grouped according to overall morphological similarity, then the tree that unites sharks and lungfishes (Fig. 1, tree 1)

would be chosen. Yet, careful morphological comparisons of extant and fossil vertebrates clearly indicates that lungfishes and tetrapods are more closely related to one another than either are to sharks²² (that is, tree 3 is correct). The apparent similarity between sharks and lungfishes is due to the retention of ancestral characteristics; these two lineages have evolved more slowly than has the lineage leading to primates.

Although molecular sequences generally evolve at a more regular 'clock-like' rate than morphological characters²³, there are many known examples of the same molecule evolving at different rates in different species²⁴. For example, baboon α -globin differs from rhesus monkey α -globin by 9 amino acids and from human α -globin by 11 amino acids, whereas the human and rhesus proteins differ by only 5 amino acids^{24,25}. Of these three species, the two monkeys are most closely related. Assuming that the three α -globins are the products of the 'same' globin gene duplicate in the three primates which diverged when the species diverged (that is, the genes are 'orthologous'¹¹), then the two monkey α -globins are most closely related; the large difference in their amino-acid sequences is most probably due to very rapid evolution of the baboon protein^{24,25}. Yet, if these three primate molecules were grouped by overall similarity, the human and rhesus α -globins would cluster together. The above distinction between overall similarity and evolutionary relationship clearly applies with equal force to molecules; therefore, the task of building trees that accurately reflect evolutionary history is often more complicated than simply clustering taxa by overall similarity.

It has long been recognized^{14,26} that overall similarity can be broken into three components (Fig. 2): (1) shared-ancestral characters, which are due to retention of traits found in the common ancestor; (2) shared-derived characters, which are due to new traits or modifications that arose along more recent lines of common descent; and (3) homoplasies (convergences, parallelisms and reversals), which are due to the same new trait or modification having been derived independently along different lineages²⁷. The concept of overall similarity consciously groups 'true' evolutionary resemblance (shared-ancestral characters and shared-derived characters) together with 'false' resemblance (homoplasy)⁴. Willi Hennig¹⁴ formalized the concept that relative closeness of evolutionary relationship

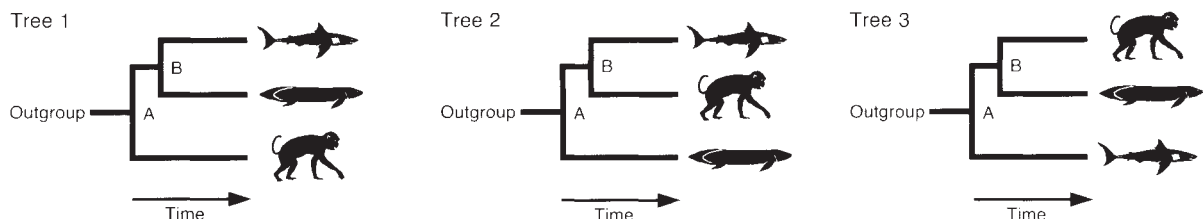


FIG. 1 Three possible trees relating sharks, lungfishes, and tetrapods. Shown here are the three possible bifurcating, rooted (node A) evolutionary trees relating sharks, lungfishes and tetrapods. In tree 1, sharks and lungfishes are depicted as most closely related; that is, they share a more recent

common ancestor (node B) than either does with tetrapods (represented here by a monkey). In tree 2, sharks and monkeys are depicted as most closely related. In tree 3, lungfishes and monkeys are depicted as most closely related. Tree 3 is thought to be correct²².

should be inferred solely on the basis of shared-derived characters. Hennig's ideas spawned the school of thought called cladistics^{4-6,15,28} from which modern parsimony analysis developed^{7-11,13}.

Powers of parsimony

The fundamental powers of parsimony analysis are that it uses derived characters to infer phylogenetic trees (Box 1), and that it can be used to tweeze apart overall similarity into its various components¹⁴ (Fig. 2). Parsimony analysis may well be the most versatile and powerful tool in evolutionary biology; but, like any power tool, its proper use requires training, practice and attentiveness.

Building phylogenetic trees. The essence^{18,29} of building parsimony trees from molecular sequences⁷⁻¹¹ is illustrated in Box 1. At its simplest, the method is a purely logical one that partitions similarities on a character-by-character basis. Alternative trees are evaluated, one character at a time, to determine how many evolutionary events they each require. By the criterion of parsimony, the 'best' or most parsimonious tree is the one that requires the fewest total events^{30,31}. At the operational level, parsimony analysis does not distinguish between shared-derived similarities and homoplasies³². The most parsimonious tree will be the correct phylogeny only if the number of shared-derived characters is high enough and the number of homoplasies low

enough. Recent studies using known phylogenies and actual molecular data have shown that parsimony analysis is generally reliable for the inference of the correct tree^{33,34}. Parsimony and other methods of phylogenetic tree-building have been reviewed recently^{9,17,19}.

It is simple to examine all possible trees for four or five taxa manually. Indeed, building trees by hand is the best way to understand the method, and is a highly recommended exercise (Box 1). But, as the number of taxa increases, the number of possible trees increases in a greater than exponential manner³⁰, and building trees quickly becomes a task best suited for a computer.

A versatile and sophisticated computer package for inferring parsimony trees, *Phylogenetic Analysis Using Parsimony* (PAUP)³², has been developed by David Swofford. Although other good parsimony programs are available (listed in refs 9 and 20), PAUP is the most generally useful package, and will be discussed here. PAUP can guarantee to find the most parsimonious trees for relatively large data sets, and has many invaluable features and options. Using various user-defined assumptions, PAUP can analyse any type of character data (such as nucleic acid sequences, protein sequences, restriction site polymorphisms, morphological characters, behavioural characters, and so on). The ability to analyse diverse data sets with the same basic method and computer

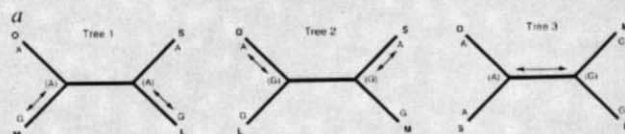
BOX 1 Building parsimony trees using nucleotide sequences

The simplest case¹⁸ for undirected parsimony analysis involves a minimum of four taxa. There are three possible unrooted bifurcating trees for four taxa, as illustrated in *a*, using shark (S), lungfish (L), monkey (M) and a non-vertebrate outgroup (O). Parsimony analysis of nucleotide sequences requires homologous (evolutionarily related) and properly aligned sequences; each nucleotide position in the sequence is considered a 'character' which can have five states (G, A, T, C or gap). Ten nucleotides from a hypothetical gene for the four taxa are aligned in *b*. Characters 1, 3 and 7 are invariant (shared-ancestral for all four taxa); invariant and highly conserved characters allow alignments and inference of homology (evolutionary relationship) between sequences. Only 'phylogenetically informative' sites (those that can be used to choose between alternative trees, in this case because two taxa have one nucleotide and two other taxa have another nucleotide) are useful in building parsimony trees¹⁸. The phylogenetically informative characters are marked with an asterisk (*). The other variable sites (characters 4 and 10) are not phylogenetically informative because all trees require the same number of substitutions.

Choosing between alternative trees: To find the most parsimonious tree, each character is evaluated on the unrooted trees to determine how few substitutions are needed to explain its observed distribution. For example, character 2 is analysed on the three unrooted trees in *a*. As indicated by the arrows, tree 1 requires a minimum of two nucleotide substitutions (one of two equally parsimonious solutions is shown, with A assumed at the nodes, and changes to G occurring along the lineages leading to M and L), tree 2 requires a minimum of two substitutions (again, only one of the parsimonious solutions is shown), and tree 3 requires only one substitution. If each character in the sequence were analysed in this manner (and the reader should do this), the events per tree would be as shown in *b*.

The optimal or 'best' tree by the parsimony criterion is the one that requires the fewest total evolutionary events^{30,31} (nucleotide substitutions, in this case); that tree is said to be the 'most parsimonious' or 'shortest' tree. The shortest tree in this hypothetical example is tree 3, because it requires 10 substitutions, whereas trees 1 and 2 each require 12. The sites that are compatible with each tree are underlined; the number of compatible sites per tree can be used in statistical tests in hopes of ruling out alternatives^{29,39}.

Rooting the tree: The tree produced by parsimony programs such as PAUP³² is 'unrooted', meaning that the ancestral node has not been identified. Rooting any tree requires knowledge or assumptions about the data set; decisions regarding rooting cannot be abdicated to the computer. Three common ways of rooting molecular trees are as follows. (1) Outgroup. One or more lineages can be included that are known to have diverged before the divergence of the taxa of interest. Ideally, the outgroup taxon



b	Taxa	Character									
		5	10								
	Shark	G	A	T	C	C	T	A	G	G	C
	Lungfish	G	G	T	C	A	C	A	T	G	T
	Monkey	G	G	T	C	A	T	A	T	C	T
	Outgroup	G	A	T	A	C	C	A	G	C	A
		*		*	*	*	*	*	*	*	*
Events per tree 1		0	2	0	1	2	2	0	2	1	2
Events per tree 2		0	2	0	1	2	1	0	2	2	2
Events per tree 3		0	1	0	1	1	2	0	1	2	2
		Total: 12									
		Total: 12									
		Total: 10									

should be a member of a closely related sistergroup to the taxa of interest, the ingroup. (2) Gene duplication. When dealing with multigene families, a known earlier gene duplicate can be used as an outgroup to root the tree¹⁰. If gene duplicates are used, a final step is needed to complete the rooted tree: gene duplications must be assigned at internal nodes in sufficient number to explain the known distribution of the duplicates within extant species. (3) Midpoint. If no information is available regarding outgroups, the root can be placed at the half-way mark, or midpoint, along the longest reconstructed lineage between two taxa³². Midpoint rooting rests on the shaky assumptions of relatively equal rates of evolution along different lineages and appropriate sampling of the lineages.

To visualize turning an unrooted tree into a hierarchical phylogeny, think of the unrooted tree as though it were made of string, and mentally tug on the ancestral node until the other lineages move into place (or make a string tree and do this manually). Tugging on the outgroup of each of the unrooted trees shown here will produce the respective rooted trees in Fig. 1.

program is helping to unite the subdisciplines of evolutionary biology.

In theory, a tree built from molecular sequences (the 'gene tree') should be an accurate reflection of the evolutionary history of those sequences. An accurate gene tree will mirror the phylogeny of the species from which the molecules were obtained if, and only if, the molecules being compared are orthologous, are not different alleles that have been retained across speciation events, and have not been the victims of gene conversion or other disruptive events after the speciation^{18,29,35}. The inference of species phylogenies from gene trees is in widespread practice today, and has yielded many important and unexpected results^{36,37}. Gene trees also inform us about the evolution of multigene families^{10,18,38}.

Objective evaluation of trees. Although the statistical properties of phylogenetic trees are not completely understood, several

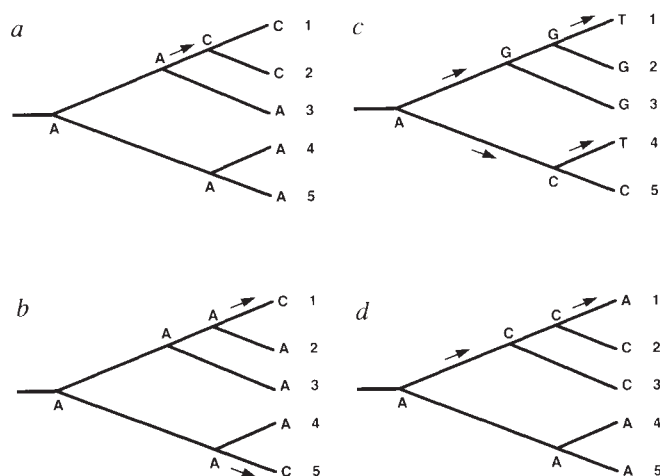


FIG. 2 The different types of similarity. The same phylogenetic tree uniting five taxa (labelled 1-5) is shown in each of the figures. The various reasons why homologous nucleotides from different taxa can be the same⁵⁶ are illustrated. These principles also apply to other unpolarized characters (that is, characters such as molecular sequences for which the ancestral state is not known *a priori*). For purposes of illustration, the ancestral states of the characters are assumed to be known: G, guanine; A, adenine; T, thymine; C, cytosine. *a*, Shared-ancestral and shared-derived characters. The Cs observed in taxa 1 and 2 are examples of shared-derived characters, whereas the As in taxa 3, 4 and 5 are shared-ancestral characters. Shared-derived characters indicate relative recentness of common ancestry, whereas shared-ancestral characters do not. Characters that are shared-derived for one level in the hierarchy may be shared-ancestral for more closely related taxa. *b*, Parallel substitutions. The Cs observed in taxa 1 and 5 are the result of nucleotide substitutions that occurred in parallel along those lineages, as indicated by the arrows. If the phylogeny were not known, this site would appear to support a tree that links taxa 1 and 5. By parsimonious reasoning on this known phylogeny, the ancestral states at the internal nodes (that is, A at each node) and the direction of substitution (a change from an ancestral A to a derived C) can be inferred from the observed sequences in the five taxa. *c*, Convergent substitutions. The Ts observed in taxa 1 and 4 are the result of convergent nucleotide substitutions because they arose from different ancestral nucleotides (G and C), and caused the sequences of taxa 1 and 4 to regain sequence similarity over time^{26,56}. (In this case, the direction of the substitutions could not be inferred from the observed sequences without further information.) The term convergence is used to mean different things in different contexts^{26,27}. For example, unrelated molecules are said to be convergent if they have gained similar functions, regardless of how this was accomplished. In common usage, the term convergence is often used in place of the term homoplasy. *d*, Reversal. The A observed in taxon 1 is the result of a change back to the ancestral state, A, which was retained in taxa 4 and 5. This reversal makes taxon 1 appear more similar to taxa 4 and 5 than it is to its closer relatives, taxa 2 and 3. (If the ancestral states were not known in this case, there would be other equally parsimonious solutions.)

statistical methods have been developed to evaluate tree hypotheses^{19,39,40}. One widely used approach is the 'bootstrap' resampling method¹⁹, which is used to evaluate the support for internal branches of a tree; this method is available as an option in PAUP. Another class of methods counts the number of sites compatible with different (usually four-taxon) phylogenies, and applies a statistical test to see if alternative trees can be ruled out^{29,39}. The ability to test alternative hypotheses objectively makes phylogenetic reconstruction a 'falsifiable' discipline⁴⁰.

Character and process analysis. Parsimony analysis attains its greatest power for the study of evolutionary pattern and process when done on rooted trees (Fig. 2), here referred to as cladistic or character analysis. Cladistic analysis can be used to reconstruct character states at the internal and ancestral nodes, and can suggest whether a given character is shared-ancestral, shared-derived, uniquely derived, or a result of homoplasy^{27,38,41-46}. This information can be used to infer modes of evolution (what happened), to calculate rates of evolution (how quickly it happened) and to detect adaptation (why it happened)^{24,45,46}. Parsimonious reasoning on phylogenies is a rigorous method in comparative biology, as is elegantly explained in two recent books^{2,3}. Although this approach is not widely used by molecular biologists, character analysis is a powerful tool for studying the evolution of the molecules themselves^{10,11,38,44-46}. For example, the most parsimonious explanation concerning introns in protein-coding genes is that they were inserted during early eukaryotic evolution⁴⁷.

Although character analysis can be done manually, *MacClade*⁴³, a visual and interactive Macintosh computer program, greatly aids in these calculations. The manual to this program provides a helpful review of parsimony and character analysis.

Pitfalls of parsimony

Parsimony analysis is simple and straightforward: therefore, its problems are relatively well understood. The problems encountered while building parsimony trees generally fall into two broad categories: (1) failure to find the shortest tree, and (2) the shortest tree not being the correct phylogeny. These pitfalls are discussed below, along with possible ways to detect and avoid them.

Many of the pitfalls may be avoided entirely through judicious choices regarding the number and type of taxa and characters collected. Before beginning a study, careful thought should be given to what scientific questions are to be addressed. Regardless of the major scientific questions, certain practical and methodological questions should be answered (see Box 2).

Failure to find the shortest tree. The goal of building parsimony trees is to find the shortest tree (or trees) that exists for a given data set under the chosen assumptions^{31,32}. This pursuit can fail for two practical reasons: too many taxa or too few phylogenetically informative sites, or a combination of both.

Too many taxa. Whenever possible, evolutionary tree programs should be used in a mode that guarantees to find all optimal trees. PAUP³² has two exact algorithms that guarantee to find all most-parsimonious trees for a limited number of taxa. The task of finding the shortest tree becomes prohibitively time-consuming as the number of taxa increases, even for the most sophisticated and rapid computers and programs. For up to 10 taxa, PAUP can do an exhaustive search of all possible trees, and produce a frequency distribution showing the tree lengths. Using a 'branch and bound' algorithm, PAUP can guarantee to find the shortest tree for a maximum of about 12 to 25 taxa, depending on the data, but will not give the distribution³². The practical upper limit for the number of taxa depends on the length and complexity of the character data, the assumptions and options chosen, the speed of the computer, and the patience of the investigator. To analyse a large number of taxa, a computer run may take hours, days or even weeks. Because it often takes longer to analyse data properly than to generate it, phylogenetic

analysis should not be tacked on to the end of a study as an afterthought. A few suggestions concerning the analysis of large data sets follow.

If the data set has too many taxa for exact algorithms, question whether all taxa are necessary. Rarely is the exact placement of every available taxon relevant to the question of interest. Sometimes taxa can be selectively omitted, provided that important conclusions are not altered significantly by the choice of omitted data⁴⁸.

Alternatively, a large set of taxa can be broken into smaller groups⁴⁹. For example, one may be interested in the branching order of certain major lineages (such as plants, animals and fungi), for which numerous sequences are available. Rather than omitting taxa, one might find the most parsimonious tree for the species within each major lineage, then define the topologies of these sub-trees to PAUP³². PAUP will treat each defined sub-tree as a single taxon while building trees that relate the lineages, thereby greatly reducing the number of alternative trees that must be considered. Moreover, PAUP will reconstruct ancestral sequences for the sub-trees which should better represent the lineages than would any of the individual sequences. Thus, the power of phylogenetic reconstruction^{10,11,42,43} can be combined with the hypothesis testing abilities of the four-way test^{29,39} to focus on the question of interest.

For some questions, neither of the above taxa reduction approaches may be appropriate. If so, parsimony programs can be run using faster 'heuristic' algorithms that attempt, but do not guarantee, to find the shortest trees. Heuristic parsimony algorithms rearrange starting trees to look for shorter ones, and can get trapped in local optima where no simple rearrangement will yield the globally most parsimonious tree. A way to escape local optima is to use many different starting trees⁵⁰; this can be done automatically in PAUP through its random addition option³².

Distance trees also can be used as starting trees in heuristic parsimony searches. Currently, most distance tree algorithms are heuristic and will analyse a large number of taxa quickly, yet will produce only one 'optimal' tree per run. With heuristic algorithms, the input order of the data can influence the branching order of the tree that is found. Therefore, the input matrix

should be reordered and the program run repeatedly to search for optimal trees²⁰. Ideally, optimal distance and parsimony trees should be found and compared; congruence of the topologies is reassuring, but does not guarantee that the correct tree for that data has been inferred. The computer software package, PHYLIP²⁰, contains programs for most distance-matrix methods, as well as some molecular parsimony programs. Taken together, PAUP and PHYLIP cover most methods for phylogenetic inference.

Too few phylogenetically informative sites. If a data set does not contain enough shared-derived, phylogenetically informative characters to resolve the branching order of all of the taxa, many equally parsimonious trees may be found. What constitutes 'enough' such characters depends on the type and quality of data; as a minimum, the data set generally should contain more phylogenetically informative characters than it does taxa. Poorly supported branches in the trees will have low bootstrap scores. Complete lack of resolution of lineages can be detected by branches of zero length. Distance methods sometimes can perform better than parsimony for data sets having few phylogenetically informative sites, because all variable sites are used in the distance calculations⁵¹.

The problem of too few phylogenetically informative characters can arise if too short a segment of DNA is sequenced, or if a region is chosen for study that is not evolving rapidly enough to answer the question at hand. It is a good idea to make analysis an ongoing part of the research strategy, because it can provide valuable information concerning adequacy of the data. The obvious solution to this problem is to collect more appropriate characters per taxa, but this is not always feasible. It may be difficult or impossible to collect enough phylogenetically informative characters to fully resolve a gene phylogeny for very closely related taxa, such as individuals within populations or species.

The data set from the well publicized human mitochondrial DNA study⁵² illustrates this point. In this extensive study, 610 nucleotides from the rapidly evolving mitochondrial control region were sequenced from 189 individuals; of the 610 characters, 201 were variable and 119 were phylogenetically informative⁵². This data set contains too many taxa for an exact search, and too few phylogenetically informative sites to allow complete resolution of the phylogeny by parsimony; indeed numerous equally parsimonious trees exist for these data^{50,52}. This example should be a cautionary tale for those who wish to study population genetics phylogenetically: the combined problems of too many taxa and too few phylogenetically informative sites are likely to plague all such studies. Indeed, the literature contains numerous examples that have not been so well publicized.

The shortest tree not being the correct phylogeny. Even if the data set appears to have plenty of phylogenetically informative sites and an exact search finds one most parsimonious tree, that tree may not accurately reflect the true evolutionary history of the taxa. There are conditions under which parsimony analysis can fail to find the correct phylogenetic tree.

Homoplasy, in its various disguises, is the ultimate trickster of parsimony. Identities or similarities due to convergence, parallel evolution, and reversals can cause historically incorrect trees to be most parsimonious. Convergence and parallel evolution are often considered indications of adaptation or positive selection, but a certain amount of homoplasy happens by chance^{27,53}.

Chance homoplasy. The more distantly related the taxa, the more likely that multiple substitutions have occurred at variable sites in the sequences, especially at synonymous sites in codons. Multiple hits obscure the phylogenetic 'signal' (the informative sites that support the true phylogeny) with 'noise' (homoplasy). A noisy data set cannot produce the correct phylogeny with any certainty⁵⁴; yet low levels of random homoplasy are unlikely to produce an incorrect tree having significant support because the events are usually scattered over the tree and cancel each other⁵³.

BOX 2 Practical and methodological questions concerning evolutionary trees

Below are examples of some basic questions that should be answered regarding phylogenetic analyses; generally this information should be presented with published phylogenies. If space permits, the complete sequence alignment should also be presented, as should any derived distance matrices used in the analyses. The primary data ought to be easily available to reviewers and readers who wish to verify results or try additional analyses; submitting the data on diskettes with the manuscript is recommended.

- What computer program and algorithm was used to build the tree(s)?
- Was the program used in a manner that guarantees to find the 'best' tree?
- If not, what measures were taken to find the best tree? (For example, how many times was the data set reordered and the program rerun?)
- What criterion was used to select the tree or trees presented?
- Were there other trees that tied for best? (If so, how many? What do they look like; that is, what are their evolutionary implications?)
- If many equally parsimonious trees were found, does it make more sense to present a consensus tree?
- How many evolutionary events (that is, nucleotide substitutions, amino-acid replacements and so on) does the best tree require?
- Are there other trees that require only a few more events?
- Can these alternative trees be ruled out statistically?
- How much support (that is, bootstrap value) is there for any given branch?
- Is the gene tree a reasonable species tree? If not, what are the possible explanations?

A quick way to check a data set for phylogenetic signal is to examine the distribution of possible trees for skewness^{7,54}. PAUP automatically produces tree distributions during exhaustive searches, and a random sample of possible trees can be produced under other search modes^{32,54}. A data set with high phylogenetic signal should have a positive skew, with the shortest tree(s) on a tail of the distribution^{7,53}.

How might phylogenetic signal be extracted from noisy DNA data? One approach is to weight transversions more heavily than transitions⁵⁵, which can be done easily in PAUP³². For protein-coding genes, third positions in codons can be omitted, and only the more conservative first and second positions analysed²⁹. For distantly related proteins, it may be more productive to analyse the amino-acid sequences¹⁶ with protein parsimony programs such as PROTPARS^{20,32}.

Directed homoplasy. More interesting cases of homoplasy are those due to adaptation rather than chance²⁷. In unusual cases, 'convergence' of amino-acid sequences can cause distantly related taxa to be pulled together on amino-acid parsimony trees^{29,45,46}. A warning sign for such taxonomically localized homoplasy is the discovery of two or more short trees that dramatically rearrange a lineage, combined with an unusually long reconstructed branch length for the 'misplaced' lineage⁴⁵. Cladistic analysis of the sequences on the 'correct' rooted phylogeny will pinpoint the homoplastic characters^{45,46}. Molecular 'convergence' or 'parallelism' generally presents itself at the protein, not DNA, sequence level^{24,27,45,46}. This highlights the need for cladistic analysis of protein sequences to study adaptive evolution of protein structure and function^{38,45}.

Although homoplasy may be a pitfall of parsimony tree build-

ing, the detection of convergence and parallel evolution is one of the major powers of parsimony analysis. Character analysis on a rooted phylogeny, regardless of how the phylogeny was constructed, is the only method by which homoplasy can be detected⁵⁶.

Conclusion

Genome evolution is a rich and complex tapestry interwoven with chromosome and gene duplications, gene conversions, mobile genetic elements, allelic diversity, hybridization and, in rare cases, sequence convergence and horizontal gene transfer. Evolutionary trees built from molecular sequences reflect these complex processes. Thus, the failure of a molecular phylogeny to reflect the phylogeny of the species perfectly should not be taken as a failure of parsimony analysis. If a gene tree conflicts with an accepted species tree, one should stop and ponder why. If the same tree is found using other genes from the same species, then the molecules are probably correct about the species phylogeny. If not, the different genes may have different evolutionary histories, which can be reconstructed through careful comparative studies. Comparative analysis of genes and proteins in a phylogenetic framework informs us about molecular evolutionary processes, and sheds light on the evolution of genomes. Until recently, the primary use of parsimony analysis has been largely limited to the study of organismal evolution: its potential to resolve questions about molecular evolution is only now being realized. □

Caro-Beth Stewart is at the Department of Biological Sciences, State University of New York at Albany, Albany, New York 12222, USA.

1. Sober, E. *Reconstructing the Past: Parsimony, Evolution, and Inference* (MIT Press, Cambridge, MA, 1988).
2. Harvey, P. H. & Pagel, M. D. *The Comparative Method in Evolutionary Biology* (Oxford Univ. Press, New York, 1991).
3. Brooks, D. R. & McLennan, D. A. *Phylogeny, Ecology, and Behavior* (University of Chicago Press, Chicago, 1991).
4. Eldredge, N. & Cracraft, J. *Phylogenetic Patterns and the Evolutionary Process* (Columbia Univ. Press, New York, 1980).
5. Wiley, E. O. *Phylogenetics: The Theory and Practice of Phylogenetic Systematics* (Wiley, New York, 1981).
6. Hull, D. L. *Science as a Process* (University of Chicago Press, Chicago, 1988).
7. Fitch, W. M. in *Cladistics* (eds Duncan, T. & Stuessy, T. F.) 221–252 (Columbia Univ. Press, New York, 1984).
8. Fitch, W. M. *Am. Nat.* **111**, 223–257 (1977).
9. Swofford, D. L. & Olsen, G. J. in *Molecular Systematics* (eds Hillis, D. M. & Moritz, C.) 411–501 (Sinauer, Sunderland, MA, 1990).
10. Dayhoff, M. O. & Eck, R. V. *Atlas of Protein Sequence and Structure* Vol. 2 (National Biomedical Research Foundation, Silver Spring, MD, 1966).
11. Fitch, W. M. *Syst. Zool.* **19**, 99–113 (1970).
12. Darwin, C. *On the Origin of Species by Means of Natural Selection* (Murray, London, 1859).
13. Hull, D. L. in *The Hierarchy of Life: Molecules and Morphology in Phylogenetic Analysis* (eds Fernholm, B., Bremer, K. & Jönvall, H.) 3–15 (Elsevier, Amsterdam, 1989).
14. Hennig, W. *Phylogenetic Systematics* (translated by Davis, D. D. & Zangerl, R.) (University of Illinois Press, Urbana, 1966).
15. Wiley, E. O., Siegel-Causey, D., Brooks, D. R. & Funk, V. A. *The Complete Cladist: A Primer of Phylogenetic Procedures* (The University of Kansas Museum of Natural History, Special Publication No. 19, 1991).
16. Doolittle, R. *Of URFs and ORFs: A Primer on how to Analyze Derived Amino Acid Sequences* (University Science Books, Mill Valley, CA, 1986).
17. Nei, M. in *Phylogenetic Analysis of DNA Sequences* (eds Miyamoto, M. M. & Cracraft, J.) 90–128 (Oxford Univ. Press, New York, 1991).
18. Li, W.-H. & Graur, D. *Fundamentals of Molecular Evolution* (Sinauer, Sunderland, MA, 1991).
19. Felsenstein, J. *A. Rev. Genet.* **22**, 521–565 (1988).
20. Felsenstein, J. *PHYLIP (Phylogeny Inference Package) Version 3.5* (Computer software package and manual distributed by the author, Dept. Genetics, University of Washington, Seattle, WA, 1993).
21. Gould, S. J. *Nat. Hist.* **92**, 14–21 (1992).
22. Colbert, E. H. & Morales, M. *Evolution of the Vertebrates* 4th edn (Wiley, New York, 1991).
23. Wilson, A. C., Carlson, S. S. & White, T. J. *A. Rev. Biochem.* **46**, 573–639 (1977).
24. Gillespie, J. *The Causes of Molecular Evolution* (Oxford Univ. Press, New York, 1991).
25. Shaw, J.-P., Marks, J., Shen, C. C. & Shen, C.-K. *J. Proc. natn. Acad. Sci. U.S.A.* **86**, 1312–1316 (1989).
26. Haas, O. & Simpson, G. G. *Proc. Am. phil. Soc.* **90**, 319–349 (1946).
27. Patterson, C. *Molec. Biol. Evol.* **5**, 603–625 (1988).
28. Nelson, G. & Platnick, N. *Systematics and Biogeography: Cladistics and Vicariance* (Columbia Univ. Press, New York, 1981).
29. Irwin, D. M. & Wilson, A. C. in *Mammalian Phylogeny* (eds Szalay, F. S., Novacek, M. J. & McKenna, M. C.) 257–276 (Springer, New York, 1992).
30. Cavalli-Sforza, L. L. & Edwards, A. W. F. *Evolution* **32**, 550–570 (1967).
31. Farris, J. S. *Syst. Zool.* **19**, 83–92 (1970).
32. Swofford, D. L. *PAUP: Phylogenetic Analysis Using Parsimony, Version 3.0.s* (Computer program and manual distributed by the Center for Biodiversity, Illinois Natural History Survey, Champaign, IL 61820, 1992).
33. Atchley, W. R. & Fitch, W. M. *Science* **254**, 554 (1991).
34. Hillis, D. M., Bull, J. J., White, M. E., Badgett, M. R. & Molineux, I. J. *Science* **255**, 589–592 (1992).
35. Doyle, J. *Syst. Bot.* **17**, 144–163 (1992).
36. Wilson, A. C., Zimmer, E. A., Prager, E. & Kocher, T. D. in *The Hierarchy of Life* (eds Fernholm, B., Bremer, K. & Jönvall, H.) 407–419 (Elsevier, Amsterdam, 1989).
37. Graur, D., Hide, W. A. & Li, W.-H. *Nature* **351**, 649–652 (1991).
38. Stewart, C.-B. *Meth. Enzym.* (in the press).
39. Li, W.-H. & Gouy, M. *Meth. Enzym.* **183**, 645–659 (1990).
40. Penny, D., Hendy, M. D. & Steel, M. A. in *Phylogenetic Analysis of DNA Sequences* (eds Miyamoto, M. M. & Cracraft, J.) 155–183 (Oxford Univ. Press, New York, 1991).
41. Swofford, D. L. & Maddison, W. P. in *Systematics, Historical Ecology, and North American Freshwater Fishes* (ed. Mayden, R. L.) 186–223 (Stanford Univ. Press, Stanford, 1992).
42. Maddison, W. P. & Maddison, D. R. *Folia Primatol.* **53**, 190–202 (1989).
43. Maddison, W. P. & Maddison, D. R. *MacClade: Analysis of Phylogeny and Character Evolution* (Sinauer, Sunderland, MA, 1992).
44. Stackhouse, J., Presnell, S. R., McGeehan, G. M., Nambiar, K. P. & Benner, S. A. *FEBS Lett.* **262**, 104–106 (1990).
45. Stewart, C.-B., Schilling, J. W. & Wilson, A. C. *Nature* **330**, 401–404 (1987).
46. Swanson, K. W., Irwin, D. M. & Wilson, A. C. *J. molec. Evol.* **33**, 418–425 (1991).
47. Palmer, J. D. & Logsdon, J. M. Jr. *Curr. Opin. Genet. Dev.* **1**, 470–477 (1991).
48. Patterson, C. in *The Hierarchy of Life* (eds Fernholm, B., Bremer, K. & Jönvall, H.) 471–488 (Elsevier, Amsterdam, 1989).
49. Sankoff, D., Cedergren, R. J. & McKay, W. *Nucleic Acids Res.* **10**, 421–431 (1982).
50. Maddison, D. R., Ruvo, M. & Swofford, D. L. *Syst. Biol.* **41**, 111–124 (1992).
51. Cornish-Bowden, A. *J. theor. Biol.* **101**, 317–319 (1983).
52. Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K. & Wilson, A. C. *Science* **253**, 1503–1507 (1991).
53. Peacock, D. & Boulter, D. *J. molec. Biol.* **95**, 513–527 (1975).
54. Hillis, D. M. in *Phylogenetic Analysis of DNA Sequences* (eds Miyamoto, M. M. & Cracraft, J.) 278–294 (Oxford Univ. Press, New York, 1991).
55. Fitch, W. M. & Ye, J. in *Phylogenetic Analysis of DNA Sequences* (eds Miyamoto, M. M. & Cracraft, J.) 147–154 (Oxford Univ. Press, New York, 1991).
56. Sneath, P. H. A. & Sokal, R. R. *Numerical Taxonomy: the Principles and Practice of Numerical Classification* (Freeman, San Francisco, 1973).

ACKNOWLEDGEMENTS. I thank R. Collura, K. Helm-Bychowski, D. Irwin, W. Messier, L. Taylor and D. Swofford for commenting on various versions of the manuscript; R. Collura for figures; D. Hillis for preprints and discussions; D. Swofford for a test version of PAUP; D. Maddison and W. Maddison for a test version of MacClade; and J. Felsenstein for PHYLIP. This paper is dedicated to the memory of Allan C. Wilson.

The chaotic obliquity of the planets

J. Laskar & P. Robutel

Astronomie et Systèmes Dynamiques, Bureau des Longitudes, 77 Avenue Denfert-Rochereau, F-75014 Paris, France

Numerical study of the global stability of the spin-axis orientation (obliquity) of the planets against secular orbital perturbations shows that all of the terrestrial planets could have experienced large, chaotic variations in obliquity at some time in the past. The obliquity of Mars is still in a large chaotic region, ranging from 0° to 60° . Mercury and Venus have been stabilized by tidal dissipation, and the Earth may have been stabilized by capture of the Moon. None of the obliquities of the terrestrial planets can therefore be considered as primordial.

THE problem of the origin of the obliquities of the planets (that is, the orientation of their spin axis) is important, because if the obliquities are primordial, they could provide some dynamical constraints on the formation of the Solar System¹⁻⁴. We have investigated how long-term perturbations by other planets affect the obliquities and precession rates of all major planets of the Solar System, and demonstrate that none of the inner planets (Mercury, Venus, the Earth and Mars) can be considered to have primordial obliquities. Any of these planets could have started with nearly zero obliquity, in a prograde state, and the chaotic behaviour of their obliquity could have driven them to their current value. If their primordial spin rate was high enough, Mercury and Venus could have undergone large and chaotic changes of obliquity from 0° to $\sim 90^\circ$, before dissipative effects ultimately drove them to their current obliquity. The present obliquity of the Earth may have been reached during a chaotic state before the capture of the Moon⁵. The obliquity of Mars is chaotic at present, with possible variations from 0° to $\sim 60^\circ$. The obliquities of the outer planets are stable, but may have gone through a similar process in an earlier stage of the Solar System's formation.

Methods

If the motion of the Solar System were quasiperiodic, the motion of precession could be modelled by equations (3) and (5) in Box 1. One can then use the argument of resonance overlap⁶ to suggest whether the motion of precession will become chaotic, or will stay regular. This can be done for the outer planets, but to obtain more accurate results, we integrated directly the full equations of precession (equations (3) and (4)), and did a frequency analysis (Box 2) of the precession solution. As the dynamical ellipticity $(C - A)/C$ is proportional⁷ to ν^2 , the pre-

cession constant α is proportional to the spin rate of the planet ν . For a given value of α , we numerically integrated the precession equations, for all values of the initial obliquity, in steps of 0.1° . These integrations use the orbital solution La90 (ref. 8) as an input and are, in general, done over 18 Myr with a step size of ~ 200 yr. We then used frequency analysis to obtain a frequency curve whose regularity tells us about the regularity of the precession and obliquity. We also recorded the maximum and minimum value reached by the obliquity during these integrations (Figs 3, 4).

The primordial spin rates of Mercury, Venus and the Earth are not known precisely, so for these planets, we did the above analysis for a wide range of α to obtain a general view of how the obliquity changed with time (Fig. 5). In Fig. 5, regular motion is represented by small dots, while the chaotic zones are larger black points. The chaotic zones would be even larger if we took into account the chaotic diffusion of the orbital solution over timescales of a billion years⁸.

Past and present obliquities

Mercury. Mercury's present spin is very low, and apparently trapped in a (2:3) spin-orbit resonance. Most probably, the

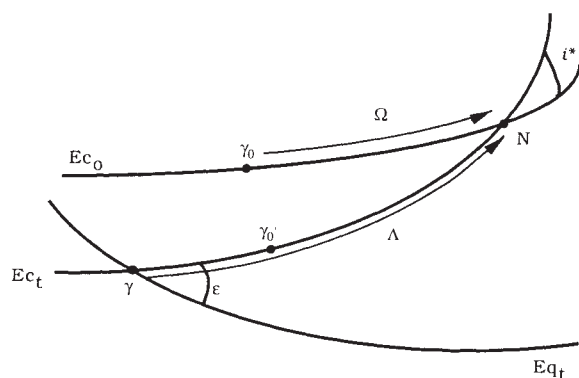


FIG. 1 Fundamental planes for the definition of precession. Eq_t and Ec_t are the mean equator and ecliptic at time t . Ec_0 is the fixed ecliptic at Julian date 2000, with equinox γ_0 . The general precession in longitude ψ is defined by $\psi = \Lambda - \Omega$. γ'_0 is defined by $\gamma'_0 N = \gamma_0 N = \Omega$; i^* is the inclination.

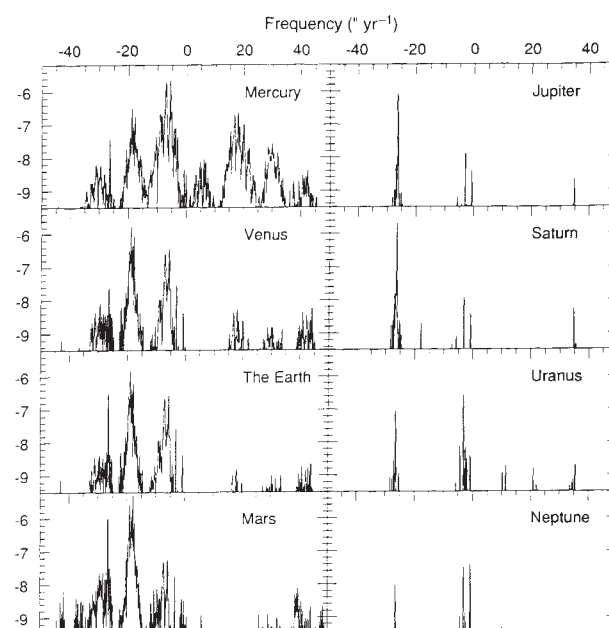


FIG. 2 Fourier spectrum (with Hanning filter) of the planetary forcing term in inclination $A(t) + iB(t)$. The logarithm of the amplitude of the Fourier coefficients is plotted against the frequency (yr^{-1}).

primordial spin rate of Mercury was much higher and was slowed down by solar tides. Extrapolating the apparent correlation of angular momentum with mass⁹, the primordial rotation period of Mercury is found to be 19 h (ref. 10), indicating a precession frequency of $\sim 127'' \text{ yr}^{-1}$. Tidal dissipation and core-mantle interactions will cause this frequency to decrease^{11,12,32}. For a rotation period smaller than 100 h, there still exists some regular motion for small obliquities, but when the spin rate of the planet decreases, for $\alpha \approx 20'' \text{ yr}^{-1}$, the obliquity enters a very large chaotic zone, extending from 0° to $\sim 100^\circ$ (Figs 3, 5a). During this period, and until tidal dissipation ultimately drives Mercury into its present state, the orientation of Mercury undergoes very large and chaotic changes from 0° to 100° within a few million years. For example, for $\alpha = 10'' \text{ yr}^{-1}$ (Fig. 3), the obliquity varies from 0 to more than 90° within a few million years for any value of the initial obliquity within this range.

Provided its primordial rotation period was smaller than 300 h (Fig. 5), Mercury must therefore have suffered large-scale chaotic behaviour during its history. At some point its orientation must have been very different from at present (with the pole facing the Sun) and this could have left some traces on the planet.

Venus. The orientation of the pole of Venus (178°) is one of the puzzling features of the Solar System^{1,13,14}. Dissipative effects of core-mantle interactions and atmospheric tides can account for some change in spin orientation, but apparently only from 90° to the present value^{13,14}. It is thus generally assumed that the retrograde rotation of Venus is primordial, which is one of the arguments supporting the stochastic accretion mechanism for the terrestrial planets^{3,4}.

As for Mercury, we investigated the global dynamics of the precession and obliquity of Venus using frequency analysis. According to Goldreich¹⁰, the primordial rotation period of Venus was ~ 13 h, which gives a precession constant of $31'' \text{ yr}^{-1}$. We started our investigation with a frequency constant of $40'' \text{ yr}^{-1}$, which reveals a regular behaviour for small obliquities, and then a large chaotic region from 50° to 90° . Tidal dissipation due to the Sun will decrease the planet's rotation speed, and consequently its precession constant. When the precession constant reaches $20'' \text{ yr}^{-1}$, corresponding to a period of ~ 20 h, the chaotic zone extends from 0° to nearly 90° (Fig. 3). In Fig. 3, we can see several well defined chaotic zones, but the frequency analysis shows that no invariant surface bounds the motion, so the orbits can diffuse in a few million years from 0° to nearly 90° . As the precession frequency continues to decrease, the shape

BOX 1 Precession and obliquity

Let $A=B<C$ be the principal moments of inertia of a planet. We assume that the axis of rotation of the planet is also its axis of maximum momentum of inertia C . The spin angular momentum is $\mathbf{H} = H\hat{\mathbf{H}}$, with $\hat{\mathbf{H}}$ a unit vector, and $H = Km\mathcal{R}^2\nu = C\nu$, where K is a coefficient depending on the internal structure of the planet ($K = 2/5$ for an homogeneous sphere), $\mathcal{R}$ its equatorial radius, m its mass, and ν its spin rate. Let $\hat{\mathbf{r}}$ be the unit vector in the direction of the Sun, and r the distance to the Sun. The torque exerted by the Sun, limited to first order in $\mathcal{R}/r$, is

$$\mathbf{L} = \frac{3GM}{r^3} \hat{\mathbf{r}} \times \mathbf{I} \hat{\mathbf{r}} \quad (1)$$

where $\mathbf{I}$ is the matrix of inertia^{23,24}, G the gravitational constant, and M the solar mass. The precession motion of the planet is given by $d\mathbf{H}/dt = \mathbf{L}$. By averaging over the time (and thus over the mean anomaly), we obtain the equations of precession; that is, the secular motion of the spin axis of the planet

$$\frac{d\mathbf{H}}{dt} = \alpha(1-e^2)^{-3/2}(\hat{\mathbf{H}} \cdot \hat{\mathbf{Z}})\hat{\mathbf{H}} \times \hat{\mathbf{Z}} \quad (2)$$

where $\hat{\mathbf{Z}}$ is the unit vector in the direction of orbital angular momentum, e the eccentricity, and where

$$\alpha = \frac{3GM}{2a^3\nu} \frac{C-A}{C}$$

with a the semi-major axis of the planet, will be called the precession constant. Let E_{co} be the orbital plane at the origin of time with equinox γ_0 , E_{ct} the orbital plane at time t , with equinox γ (Fig. 1), and N the intersection of E_{co} and E_{ct} . The precession in longitude is $\psi = \Lambda - \Omega$, where $\Omega = \gamma_0 N$ is the longitude of the node, and $\Lambda = \gamma N$. We also let γ'_0 be the point on E_{ct} such that $\gamma'_0 N = \Omega$. The coordinates of $\hat{\mathbf{H}}$ in the reference frame E_{ct} with origin γ'_0 are $(\sin \psi \sin \epsilon, \cos \psi \sin \epsilon, \cos \epsilon)$, where ϵ is the obliquity (the angle between the Equator (E_{eq}) and orbital plane (E_{ct})) at time t). After some calculus, and using the action variable $X = \cos \epsilon$ and the notations $p = \sin(i^*/2) \sin(\Omega)$, $q = \sin(i^*/2) \cos(\Omega)$ where i^* is the inclination of E_{ct} with respect to E_{co} , we obtain the equations of precession

$$\frac{d\psi}{dt} = \frac{\partial \mathcal{H}}{\partial X}, \quad \frac{dX}{dt} = -\frac{\partial \mathcal{H}}{\partial \psi} \quad (3)$$

which are related to the hamiltonian

$$\mathcal{H}(X, \psi, t) = \frac{\alpha}{2}(1-e(t)^2)^{3/2}X^2 + \sqrt{1-X^2}(A(t)\sin\psi + B(t)\cos\psi) \quad (4)$$

with

$$\begin{aligned} A(t) &= 2(\dot{q} + p(q\dot{p} - p\dot{q}))/\sqrt{1-p^2-q^2} \\ B(t) &= 2(\dot{p} - q(q\dot{p} - p\dot{q}))/\sqrt{1-p^2-q^2} \\ C(t) &= (q\dot{p} - p\dot{q}) \end{aligned}$$

The expression $A(t) + iB(t)$ and the eccentricity $e(t)$ describe the orbital motion of the planet and will be given by the La90 solution of Laskar⁸. As the orbital motion of the planets, and especially of the inner planets, is chaotic^{8,18,25-27}, a quasiperiodic approximation of $A(t) + iB(t)$ is not well suited for obtaining an accurate solution over a few million years. This is reflected in the filtered Fourier spectrum of $A(t) + iB(t)$ for the inner planets (Fig. 2) where only the main peaks can be identified as combinations of the planetary secular fundamental frequencies. A quasiperiodic approximation of $A(t) + iB(t)$ of the form

$$A(t) + iB(t) \approx \sum_{k=1}^N \alpha_k e^{i(\nu_k t + \phi_k)}$$

can still be used when looking at the outer planets or, more important, for a qualitative understanding of the behaviour of the solution. With this approximation, the hamiltonian reads

$$\begin{aligned} H &= \frac{\alpha}{2}(1-e(t)^2)^{3/2}X^2 + \sqrt{1-X^2} \\ &\times \sum_{k=1}^N \alpha_k \sin(\nu_k t + \psi + \phi_k) \end{aligned} \quad (5)$$

which is the hamiltonian of an oscillator of frequency αX , perturbed by a quasiperiodic external oscillation with several frequencies ν_k . Resonance will occur when $\psi \approx \alpha X = \alpha \cos \epsilon$ is equal to the opposite ($-\nu_k$) of one of the frequencies ν_k . When limited to a single periodic term, and with $e(t) = Cte$, this hamiltonian is integrable and is often referred to as Colombo's top²⁸. Its equilibrium points are then called Cassini states²⁹. But in the present case, $e(t)$ and $A(t) + iB(t)$ take into account the secular perturbations of the whole Solar System, modelled as a system with 15 degrees of freedom, and this hamiltonian has $1+15$ degrees of freedom. Its solutions evolve in a phase space of dimension $2+30$ (2 accounts for the ψ, X variables). In fact, the orbital solution La90 is not coupled with the precession variables, and will thus never change. The hamiltonian is thus considered to depend on time, but in a nonperiodic way. Traditional tools, such as the Poincaré surface of sections, cannot be used for the analysis of its dynamics, and we will rely on Laskar's method of frequency analysis^{8,30,31} (Box 2).

of the chaotic region changes (Fig. 5). If the obliquity of Venus reaches $\sim 90^\circ$, it could become stabilized around that value and evade the chaotic zone. Dissipative effects could then, as the planet rotation continues to slow down, drive the spin axis towards its present value of 178° . Therefore the retrograde rotation of Venus may not be primordial. If the orientation underwent large-scale chaotic behaviour during its history, the pole of Venus could have faced the Sun for extended periods, possibly causing strong changes in the climate and atmospheric circulation.

Mars. The analysis for Mars is more direct, as its rotation speed can be considered as primordial. Mars is far from the Sun and has no large satellite whose tidal interactions could have slowed

it much. Ward¹⁵ showed that the planetary secular perturbations cause the obliquity of Mars, unlike that of the Earth, to suffer large variations of $\pm 10^\circ$ around its mean value 25° . The precession constant of Mars^{16,17} ($8.26'' \text{ yr}^{-1}$) is close to some secular frequencies of the planet, and the question of resonance has been raised several times^{15,17}. Without knowledge of the precise initial conditions, no definite conclusion has been drawn.

Here we investigate the problem in a different manner, by looking at its global dynamics. First we did the frequency analysis of the obliquity of Mars over 45 Myr. We found a large chaotic zone ranging from $\sim 0^\circ$ to 60° . In Fig. 4a and b, the motion looks regular for small values of the initial obliquity. But the motion of the inner planets is chaotic^{8,18}, and the phases of the planetary secular terms are lost after ~ 100 Myr. We can thus study the problem independently of the phases. The results in Fig. 4c, d were calculated with a small change of the initial

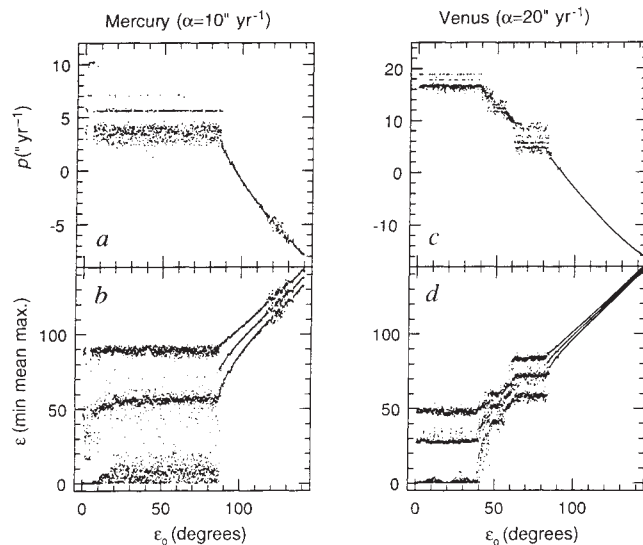


FIG. 3 Frequency analysis of the obliquity over 18 Myr for Mercury (a, b) with $\alpha = 10'' \text{ yr}^{-1}$ (rotation period ~ 255 h) and Venus (c, d) with $\alpha = 20'' \text{ yr}^{-1}$ (rotation period ~ 20 h). The precession frequency is plotted against the initial obliquity. The regularity of the frequency curve shows the regularity of the motion. For Mercury, the chaotic zone extends from 0° to 100° , and from 0° to nearly 90° for Venus. The maximum, mean and minimum values of the obliquity reached over 18 Myr are shown in b and d.

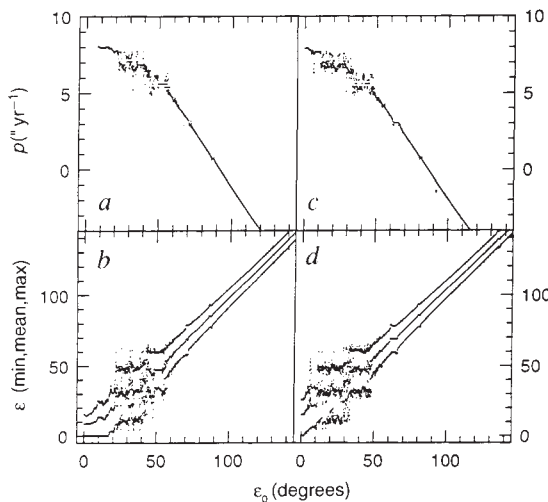


FIG. 4 Frequency analysis of the obliquity of Mars over 45 Myr. In a and b, the planetary solution is used with the present initial conditions. A large chaotic zone is visible, ranging from 0° to 60° . When the phase of the planetary solution is shifted by about -3 Myr, this large chaotic zone is even more visible (c, d). The maximum, mean and minimum values of the obliquity reached over 45 Myr are shown in b and d.

BOX 2 Frequency analysis

The numerical analysis of the fundamental frequencies follows the method of ref. 8. In that case, frequency analysis showed that for the inner planets (Mercury to Mars) the chaotic zone is relatively large, whereas for the outer planets (Jupiter to Neptune) this zone is much smaller. More generally, the method outlined here can be applied to study the stability of the solutions of a conservative dynamical system, and is based on a refined numerical search for a quasiperiodic approximation of its solutions over a finite time span^{8,30,31}. If $f(t)$ is a function with values in the complex domain obtained numerically over a finite time span $[-T, T]$, the frequency analysis algorithm will consist of a search for a quasiperiodic approximation for $f(t)$ with a finite number of periodic terms of the form

$$\tilde{f}(t) = \sum_{k=1}^N a_k e^{i\omega_k t}$$

The frequencies ω_k and complex amplitudes a_k are found by iteration. To determine the first frequency ω_1 , one searches for the maximum amplitude of $\phi(\sigma) = \langle f(t), e^{i\sigma t} \rangle$, where the scalar product of two functions $\langle f(t), g(t) \rangle$ is defined by

$$\langle f(t), g(t) \rangle = \frac{1}{2T} \int_{-T}^T f(t) \overline{g(t)} \chi(t) dt$$

and where $\chi(t)$ is a weighting function, that is, a positive function with $1/2T \int_{-T}^T \chi(t) dt = 1$, and $\overline{g}$ the complex conjugate of g . Once the first periodic term $e^{i\omega_1 t}$ is found, its complex amplitude a_1 is obtained by orthogonal projection, and the process is started again on the remaining part of the function $f_1(t) = f(t) - a_1 e^{i\omega_1 t}$. A second analysis is made to estimate the precision of the determination. In the case of an integrable hamiltonian system with n degrees of freedom, the frequency analysis of the solutions will give their quasiperiodic expansion and in particular will determine the vector $(\nu_i)_{i=1,n}$ of the fundamental frequencies of the system.

If an orbit is not regular (not quasiperiodic), in case of nearly integrable systems, the frequency analysis gives a quasiperiodic approximation to the solution which holds only locally in time. In other words, it will give us a frequency vector $(\nu_i(t))_i$ for each value of t , obtained by applying the frequency analysis algorithm over the time span $[t, t+T]$. For the analysis of the precession and obliquity, we shall only look for the precession frequency p , which is associated to the angle and action variables ψ, X . The initial phase ψ_0 is fixed. If we take some initial conditions X , we can carry out the frequency analysis for the orbits corresponding to initial conditions (ψ_0, X) (at $t=0$) over the time span $[0, T]$. We thus define the frequency map

$$F_T: \mathbf{R} \rightarrow \mathbf{R} \\ X \mapsto p$$

The regularity of the precession and obliquity can be analysed very precisely by studying the regularity of the frequency map of this problem with $1+n$ degrees of freedom. The distortions of the frequency map indicate the non-existence of invariant surfaces which may bound the obliquity variations^{30,31}. As these distortions increase, they produce a complete loss of regularity of the frequency map, which is an indication of chaotic motion³¹.

date of the orbital solution (~ 3 Myr relative to Fig. 4a, b). This should convince the reader that the chaotic zone extends from nearly 0° to 60° . Figure 4d shows that some orbits changed dramatically in less than 45 Myr.

From this, we can conclude that the obliquity of the planet Mars is chaotic, with possible variations ranging from nearly 0° to $\sim 60^\circ$. We also computed the Liapounov exponent of the obliquity of Mars for the present initial conditions¹⁹ and found a value similar to that of the planet¹⁸, $\sim 1/5$ Myr. We checked

that this value reveals the intrinsic chaotic behaviour of the obliquity of Mars by repeating the computation with a 'regular' value ($50''\text{yr}^{-1}$) of precession constant. Nevertheless, we consider the frequency analysis as more significant, as it gives the extent of the chaotic zone. Moreover, by giving the global picture of the dynamics, it shows that small changes in the initial conditions or model will not affect this result substantially.

The chaotic behaviour of the orientations of Mars could have been important in its evolution, as its obliquity could have gone

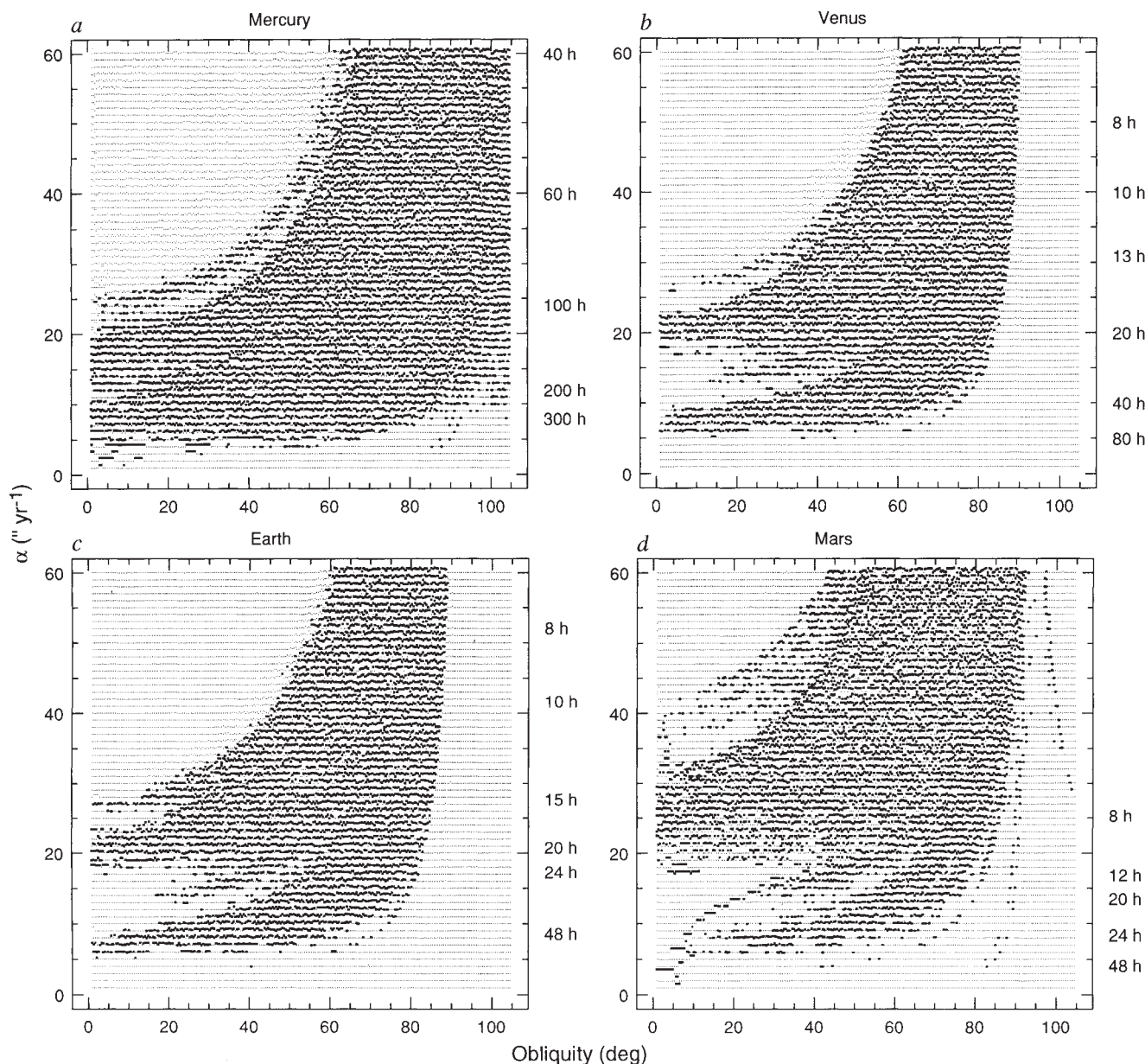


FIG. 5 The zone of large-scale chaotic behaviour for the obliquity of (a) Mercury, (b) Venus, (c) the Earth (without the Moon), (d) Mars, for a wide range of spin rate. The precession constant α is given on the left in arcseconds per year, and the estimate of the corresponding rotation period of the planet on the right, in hours. The regular solutions are represented by small dots, while large black dots represent solutions with large-scale chaotic behaviour. The chaotic motion is estimated by the diffusion of the precession frequency obtained from numerical frequency analysis over 36 Myr. For each initial condition (ϵ, α), the corresponding point is shifted vertically proportionally to the diffusion of the orbit. In the large chaotic

zones visible here, the chaotic diffusion will occur on horizontal lines (α is fixed), and the obliquity of the planet can explore horizontally all the black zone in a few million years. The extent of the chaotic zone should be even larger when one considers diffusion over longer timescales. With the Moon, one can consider that the present situation of the Earth can be represented in c roughly by the point $\epsilon = 23^\circ$, $\alpha = 55''\text{yr}^{-1}$, which is in the middle of a large zone of regular motion. Without the Moon, for a spin period ranging from about 12 h to 48 h, the obliquity of the Earth would suffer very large chaotic variations.

up to 60° during its history. The obliquity of Mars cannot be considered as primordial; the planet also could have been formed with any obliquity from 0° to 60°, and been driven to its present values only by the effect of the secular planetary perturbations.

The Earth. The equations for the precession of the Earth are complicated by the presence of the Moon; we consider the stability of its obliquity in the accompanying paper⁵. The present obliquity variations²⁰ are only $\pm 1.3^\circ$ around the mean value of 23.3°, but in the absence of the Moon, the torque exerted on the Earth would be smaller; we found, assuming a primordial rotation of 15 h, that the obliquity of the Earth would show very large variations in a chaotic zone ranging from nearly 0° to 85° (ref. 5).

The value of the primordial spin rate is controversial, so we decided to study the global stability of the obliquity for a wide range of rotation period (Fig. 5c). For a rotation period of 8 h, the increase of the equatorial bulge of the Earth would compensate for the lack of the Moon²¹, but for more generally accepted values of its spin, ranging from 12 to 48 h, most initial values of obliquity lead to large-scale chaotic variations, in many cases reaching 80–90° (Fig. 5).

For the Earth, as for the other terrestrial planets, the obliquity cannot be considered primordial. The Earth could have been formed in a large chaotic zone, and the capture of the Moon would have frozen the obliquity to its current value.

The outer planets. We also analysed the global stability of the obliquity of the outer planets (Jupiter, Saturn, Uranus, Neptune). The orbital solution of these planets behaves in a very different way from those for the inner planets. As the different spectra of $A(t) + iB(t)$ show (Fig. 2), they consist only of well isolated lines. There will be no large overlap of the different resonant terms as for the inner planets, and the solutions are essentially regular. Estimates of precession constants for the outer planets are imprecise because their structure is not well known, but are estimated²² to be well below $5'' \text{ yr}^{-1}$. As we obtain a global picture of the dynamics, we also learn how the obliquities behave for slightly different values of these constants. Practically speaking, it is difficult to destroy the stability of the obliquities until the precession constant of these planets reaches $26'' \text{ yr}^{-1}$, where it will be in resonance with the secular frequency

s_6 (ref. 8). Even then, this resonance being isolated, the motion will show large but regular oscillations. We can thus say that, as for the orbital motion, the obliquities of the outer planets are essentially stable, and should be considered as primordial.

But primordial here means that the obliquities have not changed since the Solar System was in the final state of its formation, in a state similar to its present one (at least for the outer planets). If it had, at some previous stage, been more massive than now, similar instabilities as for the inner planets could have occurred, which may have driven the obliquities of the outer planets to their present configurations.

Conclusions

None of the inner planets (Mercury, Venus, the Earth and Mars) can be considered as to have primordial obliquities, and all these planets could have been formed with a near-zero obliquity. The obliquities of all these planets could have undergone large-scale chaotic behaviour during their history. Mercury and Venus have been stabilized by dissipative effects, the Earth may have been stabilized by the capture of the Moon, and Mars is still in a large chaotic zone, ranging from 0° to 60°.

It should also be noted that planets in a retrograde state would be much more stable than planets with a prograde spin rate, as all the main forcing frequencies of the inclination are negatives (Fig. 2).

The obliquities of the outer planets (Jupiter, Saturn, Uranus and Neptune) are essentially stable, and can thus be considered as primordial, that is, with about the same value they had at the end of the formation of the Solar System. Nevertheless, chaotic behaviour of the obliquities under planetary perturbations could have occurred in an earlier stage of the formation of the Solar System, when it could have been more massive, and this point should be clarified by further studies.

The chaotic orbital motion of the inner planets^{8,18} results in large-scale chaotic behaviour of their obliquity for most initial conditions. The stability of the Earth's climate thus depends on the presence of the Moon which stabilizes its obliquity⁵, and hence the insolation variations on its surface. This should be taken into consideration when estimating the probability of finding a planet with climate stability comparable to the Earth's around a nearby star. □

Received 26 October 1992; accepted 25 January 1993.

- Harris, A. W. & Ward, W. R. *A. Rev. planet. Sci.* **10**, 61–108 (1982).
- Tremaine, S. *Icarus* **89**, 85–92 (1991).
- Dones, L. & Tremaine, S. *Science* **259**, 350–354 (1993).
- Safronov, V. S. *Evolution of the Protoplanetary Cloud and Formation of the Earth and the Planets* (Nauka, Moscow, 1969).
- Laskar, J., Joutel, F. & Robutel, P. *Nature* **361**, 615–617 (1993).
- Chirikov, B. V. *Phys. Rep.* **52**, 263 (1979).
- Lambeck, K. *The Earth's Variable Rotation: Geophysical Causes and Consequences* (Cambridge Univ. Press, 1980).
- Laskar, J. *Icarus* **88**, 266–291 (1990).
- MacDonald, G. J. F. *Rev. Geophys.* **2**, 467–541 (1964).
- Goldreich, P. & Soter, S. *Icarus* **5**, 375–389 (1966).
- Burns, J. A. *Icarus* **28**, 453–458 (1976).
- Peale, S. J. *Icarus* **28**, 459–467 (1976).
- Goldreich, P. & Peale, S. J. *Astr. J.* **75**, 273–284 (1970).
- Dobrovolskis, A. R. *Icarus* **41**, 18–35 (1980).
- Ward, W. R. *J. geophys. Res.* **79**, 3375–3386 (1974).
- Hilton, J. L. *Astr. J.* **102**, 1510–1527 (1991).
- Ward, W. R. & Rudy, D. J. *Icarus* **94**, 160–164 (1991).

- Laskar, J. *Nature* **338**, 237–238 (1989).
- Davies, M. E. *et al. Celest. Mech.* **53**, 377–379 (1992).
- Laskar, J., Joutel, F. & Boudin, F. *Astr. Astrophys.* (in the press).
- Ward, W. R. *Icarus* **50**, 444–448 (1982).
- Korycansky, D. G., Bodenheimer, P. & Pollack, J. B. *Icarus* **92**, 234–251 (1991).
- Goldreich, P. *Rev. Geophys.* **4**, 411–439 (1966).
- Murray, C. A. *Vectorial Astrometry* (Hilger, Bristol, 1983).
- Laskar, J. in *Chaos, Resonance and Collective Dynamical Phenomena in the Solar System* (ed. Ferraz-Mello, S.) 1–16 (Kluwer, Dordrecht, 1992).
- Laskar, J., Quinn, T. & Tremaine, S. *Icarus* **95**, 148–152 (1992).
- Sussman, G. & Wisdom, J. *Science* **257**, 56–62 (1992).
- Colombo, G. *Astr. J.* **71**, 891–896 (1966).
- Henrad, J. & Murigande, C. *Celest. Mech.* **40**, 345–366 (1987).
- Laskar, J., Froeschlé, C. & Celletti, A. *Physica D* **56**, 253–269 (1992).
- Laskar, J. *Physica D* (in the press).
- Peale, S. J. *Astr. J.* **79**, 722–744 (1974).

ACKNOWLEDGEMENTS. We thank P. Barge, A. Chenciner, S. Dumas, D. Gautier, M. Herman, F. Joutel, P. Y. Longaretti, S. Peale and B. Sicardy for discussions, and J. Marchand for technical assistance.

Discovery of a very bright, nearby binary millisecond pulsar

Simon Johnston*, D. R. Lorimer†, P. A. Harrison†, M. Bailes†, A. G. Lyne†, J. F. Bell‡, V. M. Kaspi§, R. N. Manchester*§, N. D'Amico¶||, L. Nicastro|| & Jin Shengzhen#

* Australia Telescope National Facility, CSIRO, PO Box 76, Epping, New South Wales 2121, Australia

† University of Manchester, Nuffield Radio Astronomy Laboratories, Jodrell Bank, Macclesfield, Cheshire SK11 9DL, UK

‡ Mount Stromlo and Siding Spring Observatory, Private Bag, Woden, ACT 2606, Australia

§ Physics Department, Princeton University, Princeton, New Jersey 08544, USA

¶ Istituto di Fisica dell'Università, via Archirafi 36, 90123 Palermo, Italy

|| Istituto di Radioastronomia del CNR, via Irnerio 46, 40126 Bologna, Italy

Department of Geophysics, Peking University, Beijing 100871, China

DURING a survey of the southern sky for millisecond pulsars, we have discovered one with by far the greatest flux density of any known millisecond pulsar, often exceeding 1 Jy at 430 MHz. The dispersion measure (integrated electron density along the line of sight) is the smallest for any known pulsar. This object, PSR J0437–4715, may be the closest known pulsar, and is several times closer than any other known millisecond pulsar. Its rotation period is 5.75 ms, and it is in a 5.7-day circular orbit with a low-mass companion. The spin-down energy flux density is the third highest known, after the Crab and Vela pulsars. These properties make possible studies of the pulsar at radio wavelengths with unprecedented detail, and indicate that it should be possible to detect both the pulsar and its companion at optical and perhaps shorter wavelengths.

The survey, described briefly in ref. 1, has been searching the southern sky for millisecond and other low-luminosity pulsars since July 1991, using the Parkes 64-m radiotelescope at 430 MHz. A very strong signal was seen in data taken on February 13, 1992 and observations taken July 11 1992 confirmed the presence of a 5.75-ms pulsar. The galactic latitude of -42° for this pulsar continues the trend of recent millisecond pulsar discoveries at high latitudes^{1–4}, confirming that the distribution of low-luminosity millisecond pulsars is largely isotropic⁵. The dispersion measure, $2.65 \text{ cm}^{-3} \text{ pc}$, is smaller than that for any other known pulsar, implying a distance⁶ from the Sun of only about 150 pc.

Subsequent timing observations have shown the pulsar to be in a 5.74-day circular orbit with a low-mass companion. These

TABLE 1 Observed and derived parameters of PSR J0437–4715

RA (J2000)	04 h 37 min 15.72 s $\pm$ 0.02 s
Dec. (J2000)	$-47^\circ 15' 8.2'' \pm 0.2''$
Galactic longitude l	253.4°
Galactic latitude, b	-42.0°
Period, P	5.7574518191 (9) ms
Epoch of period (MJD)	48825.0
Period derivative, $\dot{P}$	$1.2(6) \times 10^{-19} \text{ s s}^{-1}$
Dispersion measure, DM	$2.653(2) \text{ cm}^{-3} \text{ pc}$
Orbital period, P_b	5.7410425 (1) days
Projected semi-major axis, $a_p \sin i$	3.366675 (3) light s
Longitude of periastron, ω	$5(4)^\circ$
Eccentricity, e	$1.8(2) \times 10^{-5}$
Epoch of periastron (MJD)	48817.89 (7)
Mean flux density at 430 MHz	600 mJy
Mean flux density at 1,520 MHz	90 mJy
Mean flux density at 2,360 MHz	45 mJy
Spectral Index	-1.5
Magnetic field strength, B	$8.3 \times 10^8 \text{ G}$
Characteristic age, τ_c	$7.6 \times 10^8 \text{ yr}$
Distance, d	150 pc
Companion mass, m_c	$\sim 0.14/\sin i M_\odot$
Radio luminosity, L_{400}	13.5 mJy kpc^2

measurements have mostly been done at 430 MHz using a filterbank of $2 \times 256 \times 0.125 \text{ MHz}$ channels, and also at 1,520 MHz and 2,360 MHz using a filterbank of $2 \times 64 \times 5 \text{ MHz}$ channels. Further details of the observing systems are given in refs 1, 7. To obtain accurate arrival times for the pulses, data were sampled at $80 \mu\text{s}$ and folded to produce pulse profiles. Typical observations were 300 s in duration, giving an average signal-to-noise ratio of ~ 500 at all frequencies, although the pulsar flux density varies greatly between observations because of interstellar scintillation. The observed profiles were cross-correlated with a standard pulse profile appropriate for each radio frequency to obtain accurate pulse arrival times. A total of 112 pulse arrival times over a 120-day time interval were collected for this pulsar.

As this pulsar, PSR J0437–4715, is so close to the Earth, there is a good chance of optical detection of the pulsar's binary companion or even of optical pulses from the pulsar. To aid in such a study and to permit a more precise timing solution, an accurate position for the pulsar was obtained using the Australia Telescope Compact Array (ATCA). A 12-hour radio synthesis map of the field of PSR J0437–4715 was made from observations on 13 August 1992. Two separate frequency bands each of 128 MHz centred at 1,380 and 1,472 MHz were used, and the data later combined. The synthesized beam has a full width at half maximum of $6''$, and the r.m.s. background noise of the resultant image was about $200 \mu\text{Jy}$ per beam area. The pulsar was clearly detected with a flux density of 130 mJy, consistent with the pulsed observations.

The variation of the pulsar period with time (Fig. 1) is indicative of binary motion. The pulse arrival times were analysed in detail using the TEMPO timing program⁸, giving the pulsar and binary parameters presented in Table 1. The quoted position of the pulsar is that obtained from the ATCA observations. Quoted errors refer to the last significant digit; the r.m.s. phase residual for the least-squares fit was $15 \mu\text{s}$. The measured period derivative ($\dot{P} = 1.2 \times 10^{-19}$) is larger than that of most other millisecond pulsars and yields a characteristic age for the pulsar, $\tau_c = P/(\dot{P})$, of $7.6 \times 10^8 \text{ yr}$.

The remarkable strength of this pulsar is reflected in the mean flux density at the three observing frequencies given in Table 1, indicating a spectral index between 430 and 2,360 MHz of -1.5 , typical of most pulsars, but less steep than those of other millisecond pulsars⁹. Pulse profiles obtained at the three observing frequencies (Fig. 2) show significant emission over about three-quarters of the pulse period. The pulse profile appears to

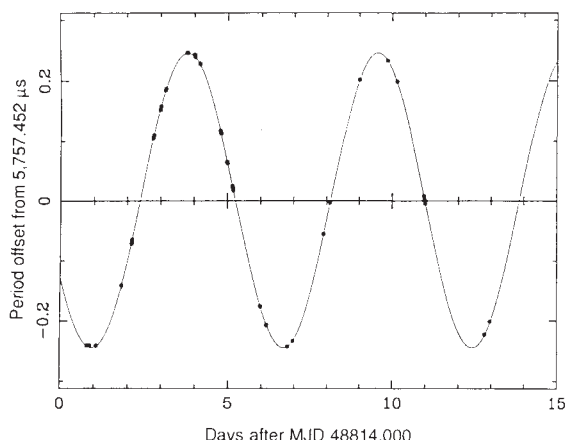


FIG. 1 The barycentric period of PSR J0437–4715 over an interval of 15 days showing the Doppler effect of the 5.7-day orbital period.

consist of a narrow central component flanked by lower-level outlying components. Unlike most other pulsars^{10,11}, the spectrum of the central emission is flatter than that of the rest, with a difference in spectral index of about 1. It is possible to detect individual pulses from this millisecond pulsar with high signal-to-noise ratio in unbroken trains. Figure 3 shows a time sequence of pulses from a 430-MHz observation. The variations from pulse to pulse are similar to those seen in ordinary long-period pulsars, providing further evidence that the emission process in millisecond pulsars is the same as that in ordinary pulsars, despite the fact that the pulsar magnetosphere is smaller in volume by a factor of about 10^6 .

The masses of the pulsar (m_p) and the companion (m_c) are constrained by the mass function

$$f(m_p, m_c) = \frac{(m_c \sin i)^3}{(m_p + m_c)^2} = \frac{4\pi^2 (a_p \sin i)^3}{G P_b^2} = 0.001243 M_\odot \quad (1)$$

where P_b is the binary period, a_p is the semi-major axis of the pulsar orbit, $M_\odot$ is the solar mass and i is the (unknown) inclination angle of the binary system to the plane of the sky. Assuming a mass of $1.4 M_\odot$ for the pulsar and $\sin i = 1$, we obtain a lower limit of $0.14 M_\odot$ for the mass of the companion. There is no evidence for any eclipse of the pulsar by the companion star, showing that it is compact.

The pulsar is a member of the slowly growing class of millisecond pulsars with low-mass companions. In the standard formation model¹², millisecond pulsars are formed when a neutron star accretes matter from an evolving companion. During this accretion phase the system may be visible as a low-mass X-ray binary. The accretion process spins up the neutron star to millisecond periods over a timescale of $\sim 10^7$ – 10^8 yr, and, when mass transfer ceases, the system consists of a millisecond pulsar in a circular orbit with a low-mass companion.

We searched for the optical companion to the pulsar on the relevant UKSTU J-survey plate. A star close to the limiting magnitude (~ 22) of the Schmidt plate lies roughly 0.7 arcsec to the west of the radio position. Until the proper motion of the pulsar is measured, we cannot be sure that this star is its optical counterpart, but it seems likely. No other stars are seen close to the radio position. At a distance of 150 pc, the distance modulus is 5.9 magnitudes providing a lower limit to the absolute magnitude of the companion of about 16 whether or not the candidate star is the binary companion. This is similar to the limit placed on the companion to PSR B1855+09 which is about 5 times more distant^{13,14}. As PSR J0437–4715 is so close, however, it should be possible to detect a companion about 25 times less luminous than the limit on PSR B1855+09. The detection of the companion could constrain models of white-dwarf temperature evolution and pulsar magnetic field decay^{14–18}.

There is also a possibility of detecting brightness variations as a function of orbital phase if the sub-pulsar point of the companion star is illuminated by high energy pulsar radiation. Such an effect is seen in PSR B1957+20, where the companion

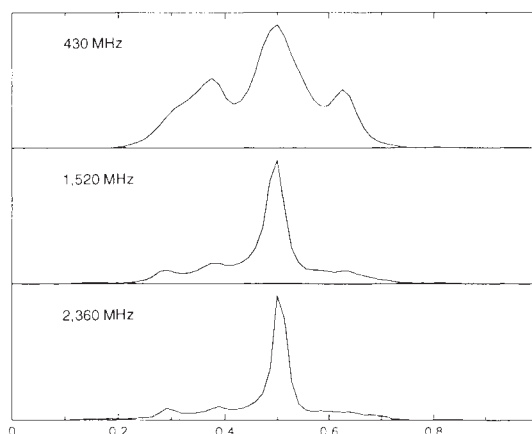


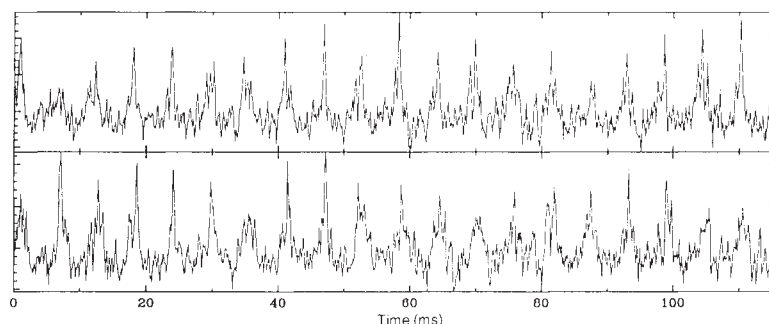
FIG. 2 Pulse profiles observed at 430 MHz, 1,520 MHz and 2,360 MHz. The whole of the pulsar period is displayed.

star varies in brightness with orbital phase by about three magnitudes¹⁹.

The proximity of the pulsar to the Earth should make it possible to determine its distance directly through the measurement of annual trigonometric parallax, the apparent shift in the pulsar position due to the motion of the Earth about the Sun. This may be detected either by using pulse timing techniques²⁰ or by very-long-baseline interferometry^{21–23}. At present, parallax measurements for pulsars are few and barely significant. In this case, however, the large flux density at high frequencies means that the expected parallax of ~ 7 mas should be measurable with high precision. The distance derived from parallax, together with the dispersion measure, could also be used to measure the local interstellar electron density and so enhance our knowledge of the pulsar distance scale based on dispersion measure⁶ and the space density of pulsars. These same astrometric techniques will also allow easy determination of the proper motion of the system and hence its velocity, important in establishing its formation history.

It has been suggested that millisecond pulsars contribute to the galactic γ -ray background, perhaps providing as much as 25% of the γ -ray flux at high galactic latitudes^{24,25}, although no millisecond pulsar has yet been detected at these energies. PSR J0437–4715 provides an excellent opportunity to detect a millisecond pulsar at γ -ray wavelengths. The pulsar rate of loss of kinetic energy is $\dot{E} = 3 \times 10^{34}$ erg s^{–1}. At the assumed distance (d) of 150 pc, this makes the spin-down energy flux density at Earth, $\dot{E}/(4\pi d^2)$, the third highest after the Crab and Vela pulsars, and several times that of PSR B1706–44, all of which were easily detected with the Gamma-Ray Observatory²⁶. Both the outer-gap and polar-cap γ -ray emission models predict that millisecond pulsars should be more efficient at producing γ -rays than high-field pulsars^{27,28}. If observations of PSR J0437–4715

FIG. 3 A 230-ms time sequence of data from PSR J0437–4715 at 430 MHz showing 40 pulse periods.



with Gamma-Ray Observatory fail to detect the pulsar, this will pose severe problems for these two γ -ray emission models. $\square$

Received 22 December 1992; accepted 27 January 1993.

1. Bailes, M. *et al. Astrophys. J.* (submitted).
2. Wolszczan, A. *Nature* **350**, 688–690 (1991).
3. Wolszczan, A. & Frail, D. A. *Nature* **355**, 145–147 (1992).
4. Nice, D. J., Taylor, J. H. & Fruchter, A. S. *Astrophys. J.* (submitted).
5. Johnston, S. & Bailes, M. *Mon. Not. R. astr. Soc.* **252**, 277–281 (1991).
6. Taylor, J. H. & Cordes, J. M. *Astrophys. J.* (submitted).
7. Johnston, S. *et al. Mon. Not. R. astr. Soc.* **255**, 401–411 (1992).
8. Taylor, J. H. & Weisberg, J. M. *Astrophys. J.* **345**, 434–450 (1989).
9. Foster, R. S., Fairhead, L. & Backer, D. C. *Astrophys. J.* **378**, 687–695 (1991).
10. Rankin, J. M. *Astrophys. J.* **274**, 333–358 (1983).
11. Lyne, A. G. & Manchester, R. N. *Mon. Not. R. astr. Soc.* **234**, 477–508 (1988).
12. Bhattacharya, D. & van den Heuvel, E. P. J. *Phys. Rep.* **203**, 1–124 (1991).
13. Callanan, P. J. *et al. Mon. Not. R. astr. Soc.* **238**, 25p–28p (1989).
14. Kulkarni, S. R., Djorgovski, S. & Klemola, A. R. *Astrophys. J.* **367**, 221–227 (1991).
15. Kulkarni, S. R. *Astrophys. J.* **306**, L85–L89 (1986).
16. Koester, D., Chanmugam, G. & Reimers, D. *Astrophys. J.* **395**, L107–L110 (1992).
17. Shapiro, S. L. & Teukolsky, S. A. *Black Holes, White Dwarfs and Neutron Stars. The Physics of Compact Objects* (Wiley-Interscience, New York, 1983).
18. Iben, I. & Tutukov, A. V. *Astrophys. J.* **282**, 615–630 (1984).
19. Fruchter, A. S., Gunn, J. E., Lauer, T. R. & Dressler, A. *Nature* **334**, 686–689 (1988).
20. Ryba, M. F. & Taylor, J. H. *Astrophys. J.* **371**, 739–748 (1991).
21. Salter, M. J., Lyne, A. G. & Anderson, B. *Nature* **280**, 477–478 (1979).
22. Gwinn, C. R., Taylor, J. H., Weisberg, J. M. & Rawley, L. A. *Astr. J.* **91**, 338–342 (1986).
23. Bailes, M., Manchester, R. N., Kesteven, M. J., Norris, R. P. & Reynolds, J. E. *Nature* **343**, 240–241 (1990).
24. Bhattacharya, D. & Srinivasan, G. *J. Astrophys. Astr.* **12**, 17–25 (1991).
25. Bailes, M. & Kniffen, D. A. *Astrophys. J.* **391**, 659–666 (1992).
26. Thompson, D. J. *et al. Nature* **359**, 615–616 (1992).
27. Harding, A. K. *Astrophys. J.* **245**, 267 (1981).
28. Cheng, K. S., Ho, C. & Ruderman, M. *Astrophys. J.* **300**, 500–521 (1986).

Stabilization of the Earth's obliquity by the Moon

J. Laskar, F. Joutel & P. Robutel

Astronomie et Systèmes Dynamiques, Bureau des Longitudes,
77 Avenue Denfert-Rochereau, F75014 Paris, France

ACCORDING to Milankovitch theory^{1,2}, the ice ages are related to variations of insolation in northern latitudes resulting from changes in the Earth's orbital and orientation parameters (precession, eccentricity and obliquity). Here we investigate the stability of the Earth's orientation for all possible values of the initial obliquity, by integrating the equations of precession of the Earth. We find a large chaotic zone which extends from 60° to 90° in obliquity. In its present state, the Earth avoids this chaotic zone and its obliquity is essentially stable, exhibiting only small variations of $\pm 1.3^\circ$ around the mean value of 23.3° . But if the Moon were not present, the torque exerted on the Earth would be smaller, and the chaotic zone would then extend from nearly 0° up to about 85°. Thus, had the planet not acquired the Moon, large variations in obliquity resulting from its chaotic behaviour might have driven dramatic changes in climate. In this sense one might consider the Moon to act as a potential climate regulator for the Earth.

Ward³ suggested that, in the absence of the Moon and on the basis of its present spin rate, the Earth might have exhibited quasiperiodic variations of about $\pm 10^\circ$ in obliquity. But he also suggested that without the Moon, the Earth would have had a faster spin rate, leading to greater rotational flattening which would have compensated for the lack of a lunar torque and reduced the obliquity variations to their present value. This conclusion requires that the primordial rotation period rate of the Earth should be less than 8 h. We find that for a slower primordial spin, and over a wide range of initial values, the obliquity variations will be chaotic, with variations much larger than predicted by Ward.

Our model and method of analysis are described in Boxes 1 and 2. We investigate the dynamics of the Earth's obliquity ε

BOX 1 Equations of precession

The general precession in longitude, ψ , and the obliquity at a given time, ε , are determined by the motions of the equatorial and ecliptic pole. The precession equations are written in terms of the action variable $X = \cos \varepsilon$ and the associated angle variable ψ . Let us denote two new variables by $p = \sin(i^*/2) \sin(\Omega)$, $q = \sin(i^*/2) \cos(\Omega)$ where i^* is the inclination of the Earth with respect to a fixed ecliptic, and Ω the longitude of the node. The equations of precession^{7,9,10} can be written

$$\frac{d\psi}{dt} = T(X, t) - \frac{X}{\sqrt{1-X^2}} (A(t) \sin \psi + B(t) \cos \psi) \quad (1)$$

$$\frac{dX}{dt} = -\sqrt{1-X^2} (B(t) \sin \psi + A(t) \cos \psi) \quad (2)$$

with

$$T(X, t) = c_1 X + c_2 (2X^2 - 1) / (1 - X^2) + c_3 (6X^2 - 1) + c_4 S_0 X - 2C(t) - p_g$$

and

$$A(t) = 2(\dot{q} + p(q\dot{p} - p\dot{q})) / \sqrt{1-p^2-q^2}$$

$$B(t) = 2(\dot{p} - q(q\dot{p} - p\dot{q})) / \sqrt{1-p^2-q^2}$$

$$C(t) = (q\dot{p} - p\dot{q})$$

The coefficients c_1 to c_4 and the geodetic precession p_g depend on the orbital parameters of the Moon and the Sun ($c_1 = 37.526603'' \text{ yr}^{-1}$, $c_2 = -0.001565'' \text{ yr}^{-1}$, $c_3 = 0.000083'' \text{ yr}^{-1}$, $c_4 = 34.818618'' \text{ yr}^{-1}$, $p_g = 0.019188'' \text{ yr}^{-1}$). We also have $S_0 = \frac{1}{2}(1-e^2)^{-3/2} - 0.522 \times 10^{-6}$, where e is the eccentricity of the Earth¹⁰. The initial conditions for the Earth are $\psi(t=0) = 50.290966'' \text{ yr}^{-1}$, $\varepsilon(t=0) = 23^\circ 26' 21.448''$. These equations can be associated with the hamiltonian depending on time

$$H(X, \psi, t) = \mathcal{T}(X, t) + \sqrt{1-X^2} (A(t) \sin \psi + B(t) \cos \psi) \quad (3)$$

where

$$\mathcal{T}(X, t) = \frac{1}{2}(c_1 + S_0 c_4) X^2 - c_2 X \sqrt{1-X^2} + c_3 (2X^3 - X) - (2C(t) + p_g) X$$

To understand the dynamics of this hamiltonian, its small terms (c_2 , c_3 , p_g , $C(t)$) can be neglected, although they will be taken into account during the numerical computations. We shall also neglect the eccentricity of the Earth and the Moon, as well as the inclination of the Moon. In this case, $c_1 = (C-A)/C \times 3n_m^2 m_M / 2\nu$, $S_0 = \frac{1}{2}$, $c_4 = (C-A)/C \times 3n_\odot^2 M_\odot / \nu$, where ν is the angular velocity of the Earth, $(C-A)/C$ its dynamical ellipticity (which is proportional to ν^2), n_m and m_M , the mean motion and mass of the Moon, and $n_\odot$ and $M_\odot$ the same quantities for the Sun. The hamiltonian then reduces to

$$H(X, \psi, t) = \frac{1}{2} \alpha X^2 + \sqrt{1-X^2} (A(t) \sin \psi + B(t) \cos \psi) \quad (4)$$

with $\alpha = c_1 + S_0 c_4$. The expression $A(t) + iB(t)$ (as well as the eccentricity e of the orbit of the Earth) will be given by the La90 solution of Laskar⁴.

($= \cos^{-1} X$ in Box 1) by integrating the equations of precession ((1) and (2)) over 18 Myr for all values of the initial obliquity ε_0 , in steps of 0.1° , from 0° to 170° . The perturbation effect of the Solar System is taken into account by using the secular La90 solution for the Earth⁴. We used the frequency analysis method⁴⁻⁶ for the analysis of the obliquity and precession (Box 2). Briefly, this method consists of defining for each orbit a frequency vector by a refined Fourier analysis over a finite time span. The regularity of the orbits can then be analysed very precisely by studying the regularity of the frequency application which links the action-like variable (ε) to the frequencies. In Fig. 2a, the precession frequency p_t is plotted against the initial obliquity ε_0 , for a fixed value of the initial precession angle ψ_0 . From $\varepsilon_0 = 0$ to $\varepsilon_0 = 60^\circ$, the frequency curve is very regular, and reflects the regular behaviour of the solution. A large chaotic zone then extends from 60° to 90° . For higher values of obliquity, the precession frequency becomes negative; the effects of possible resonances are much smaller, and the motion is again very regular. In Fig. 2b, for each integration, the minimum, maximum

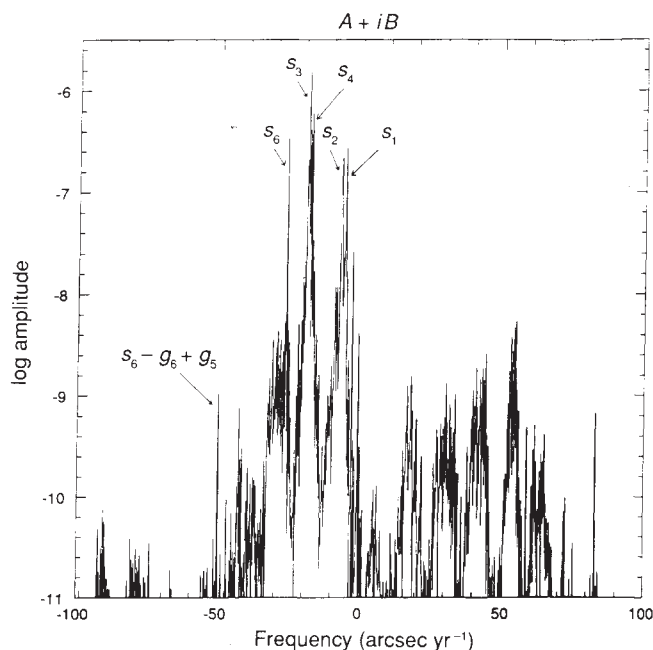


FIG. 1 Fourier spectrum of $A(t) + iB(t)$ over 17 Myr. The main secular frequencies of the Solar System can be identified, as well as the small isolated term $s_6 - g_6 + g_5$.

and mean values of the obliquity reached during the 18 Myr of the integration are given. If, for example, the obliquity goes to 60° , then in a few million years it can reach 90° , because of the planetary secular perturbations alone.

In earlier calculations⁷, we suppressed the Moon, keeping the

present values for the Earth's parameters. Because the torque exerted on the Earth is reduced ($c_1 = c_2 = c_3 = 0$), the precession frequency goes down to $\sim 15.6'' \text{ yr}^{-1}$, which is close to the value of the leading frequencies of $A(t) + iB(t)$ but with the opposite sign (Box 2). As predicted by Ward³, this led to large variations of the obliquity (from 15° to 32° in 1 Myr).

Here we look at the possible dynamics of the very early Earth, on the hypothesis that the Moon was not present at the time. If so, because of tidal dissipation, the rotation of the Earth would have been faster, as shown by geological records⁸ which give $\nu \approx 1.22 \nu_p$ for -2.5 Gyr (where ν_p is the present rotation velocity of the Earth). We analysed the obliquity variations for this rate of rotation (Fig. 3). In this case, the chaotic region becomes very large, extending from 0° to about 80° . Even if the initial obliquity is small, it can exceed 50° in a few million years. Furthermore, the frequency analysis shows the possibility of diffusion up to about 80° , although this diffusion may be slow.

For a higher rotation speed of $\nu = 1.6 \nu_p$, which may have existed at -4.5 Gyr, a large resonant zone appears, corresponding to the node secular mean rate of Jupiter and Saturn ($s_6 = -26.3302'' \text{ yr}^{-1}$) (Fig. 4). For zero initial obliquity, we find variations of $\sim 10^\circ$, but as soon as the initial obliquity reaches 4° , the motion can enter the chaotic zone surrounding the resonant island, and show variations of more than 30° in a few million years, with the possibility of further diffusion up to more than 85° .

Thus, we find that even if the initial obliquity of the Earth was very small, resonances or chaotic behaviour could have raised it to 50° in a few million years, with mean value 20° – 30° . If the Moon was then captured, the precession frequency would have increased suddenly, and the motion would have become regular. The value of the obliquity would be frozen to its current value and thereafter suffer only small oscillations, until tidal dissipation³ ultimately drives the Earth into the large chaotic zone.

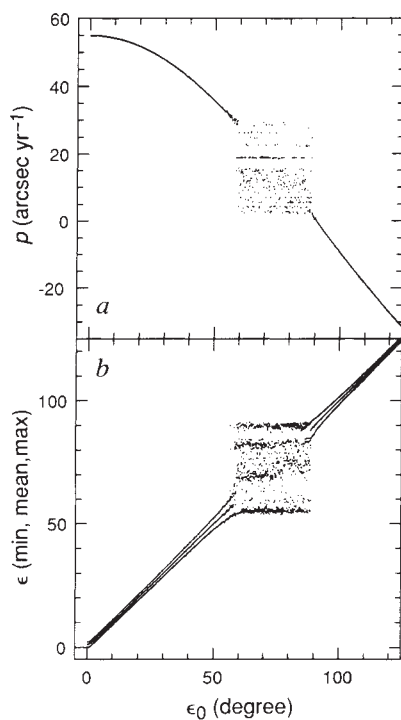


FIG. 2 Frequency analysis over 18 Myr of the precession of the Earth under lunar and solar torque for all values of the initial obliquity (ϵ_0). *a*, A large chaotic zone exists from about 60° to 90° . *b*, Maximum, mean and minimum obliquity reached during 18 Myr.

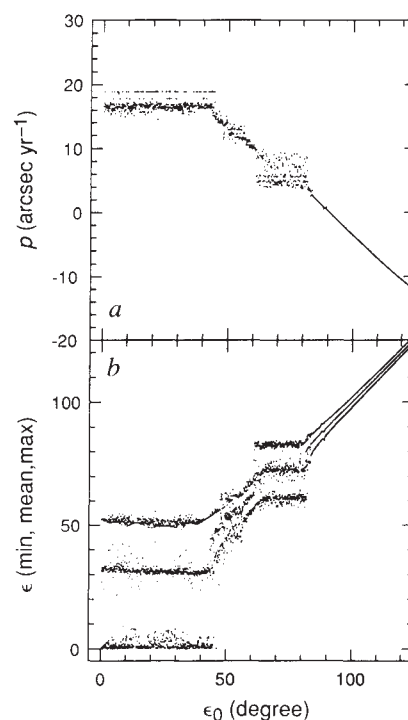


FIG. 3 Without the Moon, the chaotic zone revealed by the frequency analysis over 18 Myr (*a*) extends from 0° to $\sim 85^\circ$, for a rotation velocity of the Earth $\nu = 1.22 \nu_p$. *b*, Maximum, mean and minimum obliquity reached during 18 Myr.

BOX 2 Analysis of resonances

The solution of the orbital motion of the Earth being chaotic^{4,11-14}, a quasiperiodic approximation of $A(t) + iB(t)$ in equation (4), is not well suited for obtaining accurate solutions over a few million years, but will be useful for a qualitative understanding of the behaviour of the solution. In the Fourier spectrum of $A(t) + iB(t)$ (Fig. 1) we can observe peaks which are identified as the main planetary secular frequencies in inclination. Around each of these, several secondary peaks appear, which largely reflect the non-regular behaviour of the solution. A frequency analysis⁴⁻⁶ of $A(t) + iB(t)$ can be done in order to find a quasiperiodic approximation of this function over a few million years of the form

$$A(t) + iB(t) \approx \sum_{k=1}^N \alpha_k e^{i(\nu_k t + \phi_k)}$$

With this approximation, the hamiltonian now reads

$$H = \frac{1}{2} \alpha X^2 + \sqrt{1 - X^2} \sum_{k=1}^N \alpha_k \sin(\nu_k t + \psi + \phi_k) \quad (5)$$

which is the hamiltonian of an oscillator of frequency αX , perturbed by a quasiperiodic external oscillation with several frequencies ν_k . Resonance will occur when $\dot{\psi} \approx \alpha X = \alpha \cos \varepsilon$ is equal to the opposite ($-\nu_k$) of one of the frequencies ν_k .

The value for the mean precession speed of the Earth over 18 Myr is $p_r = 50.4712'' \text{ yr}^{-1}$. This value is very close to the opposite of a small term due to the opposite of a small term due to the perturbations of Jupiter and Saturn $s_6 - g_6 + g_5 = -50.3021'' \text{ yr}^{-1}$. The passage through resonance could occur during an ice age, and could lead to an increase of 0.5° in the obliquity variations⁷. Nevertheless, the Earth is far from the main planetary resonances, the closest being $s_6 = -26.3302'' \text{ yr}^{-1}$. With the current value of α , we estimate that this resonance is reached for an obliquity of about 60° .

It can thus be claimed that the Moon is a climate regulator for the Earth. If it were not present, or if it were much smaller, for many values of the Earth primordial spin rate, the obliquity of the Earth would be chaotic with very large variations, reaching more than 50° in a few million years and even, in the long term,

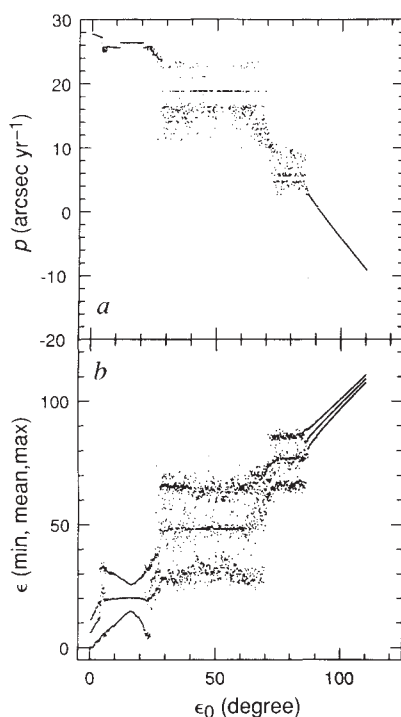


FIG. 4 For $\nu = 1.6 \nu_p$, the large chaotic zone extends from nearly 0° to $\sim 85^\circ$. a, In the region of low obliquity, there exists a large island corresponding to resonances with the secular frequency $s_6 = -26.3302 \text{ arcsec yr}^{-1}$ of the node of Jupiter and Saturn. This region is also visible in (b).

more than 85° . This would probably have drastically changed the climate on the Earth. $\square$

Received 21 September 1992; accepted 18 January 1993.

1. Imbrie, J. *Icarus* **50**, 408-422 (1982).
2. Berger, A., Imbrie, J., Hays, J., Kukla, G. & Saltzman, B. (eds) *Milankovitch and Climate* (Reidel, Dordrecht, 1984).
3. Ward, W. R. *Icarus* **50**, 444-448 (1982).
4. Laskar, J. *Icarus* **88**, 266-291 (1990).
5. Laskar, J., Froeschlé, C. & Celletti, A. *Physica D* **56**, 253-269 (1992).
6. Laskar, J. *Physica D* (in the press).
7. Laskar, J., Joutel, F. & Boudin, F. *Astr. Astrophys.* (in the press).
8. Williams, G. E. *J. Phys. Earth* **38**, 475-491 (1990).
9. Kinoshita, H. *Celest. Mech.* **15**, 277-326 (1977).
10. Laskar, J. *Astr. Astrophys.* **157**, 59-70 (1986).
11. Laskar, J. *Nature* **338**, 237-238 (1989).
12. Laskar, J. in *Chaos, Resonance and Collective Dynamical Phenomena in the Solar System* (ed. Ferraz-Mello, S.) 1-16 (Kluwer, Dordrecht, 1992).
13. Laskar, J., Quinn, T. & Tremaine, S. *Icarus* **95**, 148 (1992).
14. Sussman, G. & Wisdom, J. *Science* **257**, 56-62 (1992).

Quantum interference in a mesoscopic superconducting loop

V. V. Moshchalkov, L. Gielen, M. Dhallé,
C. Van Haesendonck & Y. Bruynseraede

Laboratorium voor Vaste Stof-Fysika en Magnetisme,
Katholieke Universiteit Leuven, B-3001 Leuven, Belgium

THE classical superconducting quantum interference device (SQUID) is based on the Josephson effect¹, and usually consists of a macroscopic superconducting loop with two artificial weak links (Josephson junctions) through which the supercurrent passes by quantum-mechanical tunnelling. Fink *et al.*² proposed a new type of SQUID based on a homogeneous mesoscopic superconducting loop, in which interference between the supercurrents passing through the two halves of the ring results in a critical current that varies with the applied magnetic field in an oscillatory manner. Here we describe the experimental observation of these oscillations in a mesoscopic superconducting aluminum loop without artificial weak links. In this new type of quantum interferometer, 'weak-link' regions with a strongly reduced superconducting order parameter appear periodically at half-integer magnetic flux quanta owing to the interplay between the shielding and transport currents in the loop.

Quantum interference in superconductors resembles the classical interference of coherent light waves. In a superconductor, the condensate of superconducting electrons can be described by the complex wave function $\psi = |\psi|e^{i\varphi}$ ('superconducting order parameter'), where the square of the wave amplitude, $|\psi|^2$, is proportional to the density of Cooper pairs. Both the amplitude $|\psi|$ and the phase φ will show pronounced spatial variations in the presence of a magnetic field $\mathbf{H}$ or a transport current I_t . In the two-slit optical interference experiment one mixes two light waves which have propagated along different paths. The phase difference between the optical waves is tuned by adapting the optical path length. In a similar way, the interference of two branching superconducting currents in a loop with two Josephson weak links (Fig. 1a) may be tuned by varying the magnetic flux Φ threading the loop. In this so-called SQUID configuration an applied magnetic field $\mathbf{H}$ perpendicular to the loop plane produces an oscillatory dependence of the critical current I_c :

$$I_c = I_0 \cos\left(\pi \frac{\Phi}{\Phi_0}\right) \quad (1)$$

where I_0 is the amplitude of the critical current oscillations and the flux quantum $\Phi_0 = (hc)/(2e) = 2.07 \times 10^{-7} \text{ G cm}^2$.

In classical SQUIDs, quantum interference requires the presence of two artificially prepared Josephson weak links (1 and 2 in Fig. 1a). Quantum interference of two supercurrents can also occur, however, in a homogeneous mesoscopic loop^{2,3} (see Fig. 1b) where the supercurrent flows from node N to node N' through two branches (1) ($I_{NN'1}$) and (2) ($I_{NN'2}$). The mesoscopic regime will be reached as soon as the loop circumference becomes comparable the Ginzburg-Landau (GL) coherence length $\xi(T)$ which diverges near the superconducting transition temperature T_c . The theoretical model of Fink *et al.*^{2,3} also assumes that both the width w and the thickness t of the loop structure are much smaller than the coherence length $\xi(T)$.

In the framework of the linearized GL theory, the supercurrents $I_{NN'1}$ and $I_{NN'2}$ are given by²

$$I_{NN'1} = \frac{T_L}{2} \sin\left(\beta - \alpha - \pi \frac{\Phi}{\Phi_0}\right) \quad (2)$$

$$I_{NN'2} = \frac{T_L}{2} \sin\left(\beta - \alpha + \pi \frac{\Phi}{\Phi_0}\right) \quad (3)$$

where α and β are the phases of the superconducting order parameter at nodes N and N' , respectively. The transmittance of a branch of the loop $T_L/2 = (e\hbar|\psi|^2)/(m\xi \sin(\pi r/\xi))$ (ref. 2) plays the role of the tunnelling probability in a conventional Josephson junction. The derivation of T_L is analogous to the impedance matching for light waves propagating through different transparent media (for details, see ref. 2). The interference of the currents flowing through the two branches of the loop with radius r produces an oscillatory dependence of the total current $I_{NN'} = I_{NN'1} + I_{NN'2}$

$$I_{NN'} = T_L \sin(\beta - \alpha) \cos\left(\pi \frac{\Phi}{\Phi_0}\right) \quad (4)$$

The current amplitude in equation (4) is determined by the loop transmittance T_L and by the phase difference $\Delta\varphi = \beta - \alpha$ (as in the first Josephson equation $I \propto \sin \Delta\varphi$).

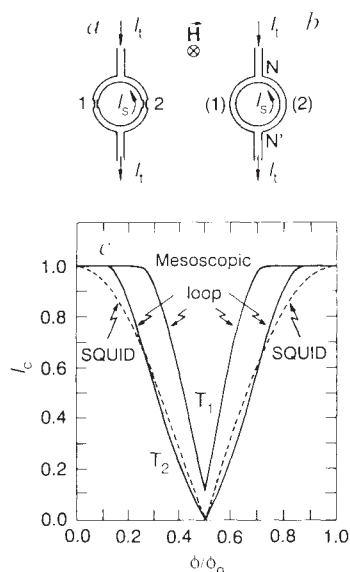


FIG. 1 a, Conventional SQUID with two artificial Josephson weak links 1 and 2. b, Homogeneous mesoscopic superconducting loop without artificial weak links. c, Normalized critical current I_c of a homogeneous mesoscopic superconducting loop (solid lines) as a function of the normalized magnetic flux Φ/Φ_0 for a ratio of the loop radius to the superconducting coherence length $r/\xi(T) = 0.25$ and 0.5 , corresponding to temperatures T_2 and T_1 ($T_2 > T_1$), respectively. For comparison, the oscillatory behaviour of the normalized critical current of a conventional SQUID is shown by the dashed line (adapted from ref. 3).

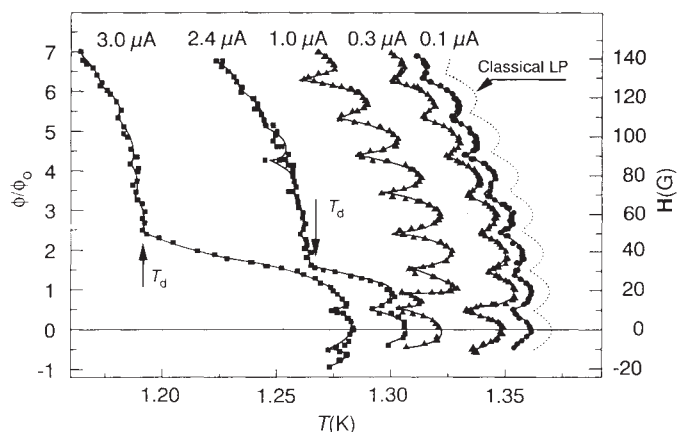


FIG. 2 Measured phase boundary in the H - T plane of a mesoscopic $1 \mu\text{m} \times 1 \mu\text{m}$ Al loop for different transport currents. The data points, which define the boundary, have been obtained by measuring the position of the midpoint of the normal to superconducting transition for different magnetic fields. The dotted line shows the classical Little-Parks phase boundary, calculated using the Tinkham model¹⁰, with $r = 0.57 \mu\text{m}$, $w = 0.17 \mu\text{m}$, $T_c(H=0) = 1.363 \text{ K}$, and $\lambda_L H_{c0} = 5.4 \times 10^{-3} \text{ G cm}$, where λ_L and H_{c0} are the London penetration depth and the thermodynamic critical field for Al. The full lines are a guide to the eye.

The critical current I_c for the homogeneous loop is obtained by taking $\sin(\beta - \alpha) = 1$ in equation (4)

$$I_c = T_L \cos\left(\pi \frac{\Phi}{\Phi_0}\right) \quad (5)$$

Equation (5) implies that the critical current of a mesoscopic superconducting loop without artificial Josephson weak links oscillates with the applied magnetic field in the same way as it does in a classical SQUID with extrinsic weak links (equation (1)). Equation (5) has been derived in the framework of the linearized GL theory which is used to describe superconductivity in a homogeneous wire with thickness $t \ll \xi(T)$ and width $w \ll \xi(T)$. When this quasi one-dimensional approach^{4,5} is extended to include the nonlinear GL term^{3,6}, the oscillations of the critical current still persist, but plateaus are introduced around integer flux quanta (Fig. 1c).

Combining electron beam lithography and lift-off techniques, we prepared square aluminium loops with a side of $1 \mu\text{m}$ and a linewidth $w = 0.15 \mu\text{m}$. The $0.025 \mu\text{m}$ thick Al lines are obtained by thermal evaporation of high-purity Al (99.9999%) in a reduced helium atmosphere ($p \approx 10^{-3} \text{ Torr}$), which, according to our experience, produces smooth and homogeneous metallic films. The electrical transport measurements are done in a temperature-stabilized helium-3 cryostat. The loops typically have a normal-state resistance $R_n \approx 20 \Omega$. Small variations of the loop resistance can be detected using a commercial four-terminal a.c. resistance bridge (Linear Research 400).

An unambiguous experimental verification of the predictions made by Fink *et al.*^{2,3} is only possible when (1) the dimensions t, w are much less than $\xi(T)$ and (2) the presence of extrinsic weak links in the superconducting loops can be avoided. The first condition is fulfilled as the superconducting coherence length in our Al structures is larger than $0.6 \mu\text{m}$ for $T > 1.3 \text{ K}$ (ref. 7). Concerning the second condition, severe cracks at the grain boundaries in the Al lines, or discontinuities caused by a local failure of the lift-off process may induce a weak Josephson coupling between different line segments. We are confident that with our preparation method the accidental formation of such weak links in the Al loops is unlikely. The resistivity $\rho \approx 4.0 \times 10^{-6} \Omega \text{ cm}$ of the Al lines and the resistance ratio $R(300 \text{ K})/R(4.2 \text{ K}) \approx 2.1$ are comparable to the values measured for the much larger Al contact pads and indicate a pronounced

metallic character. Moreover, detailed atomic force microscopy and scanning electron microscopy studies of the Al films confirm that their surface is smooth and continuous.

The quasi one-dimensional approach of Fink *et al.*^{2,3} imposes additional constraints on the temperature interval where the $I_c(H)$ oscillations (equation (5)) can be observed. To determine this temperature interval we will first focus on the effect of the transport current I_t on the $T_c(H)$ phase boundary^{7,8}. This boundary (see Fig. 2) is obtained by measuring the midpoint of the normal to superconducting resistive transitions for different magnetic fields (for experimental details see ref. 8). Three current regimes can be distinguished. For transport currents $I_t < 0.3 \mu\text{A}$, corresponding to a current density $J_t \approx 8 \times 10^3 \text{ A cm}^{-2}$, the phase boundary shows maximum T_c values at integer flux quanta $\Phi/\Phi_0 = n$, typical for the classical Little-Parks (LP) oscillations⁹ caused by the flux quantization. For magnetic fields corresponding to half-integer flux quanta, the induced screening currents reach a maximum, resulting in a maximum depression of T_c . The experimental $T_c(H)$ dependence is in good agreement with the calculated behaviour¹⁰ (dotted line in Fig. 2). For intermediate currents $0.3 \mu\text{A} \leq I_t < 2.4 \mu\text{A}$ in the mesoscopic loop, anomalous LP oscillations are observed with a pronounced T_c enhancement at low flux quanta $\Phi/\Phi_0 < 2$. Finally, for currents $I_t \geq 2.4 \mu\text{A}$ the $T_c(H)$ oscillations decrease markedly, and a well developed kink appears at a temperature $T = T_d$.

The kink in the $T_c(H)$ curve at $T = T_d$ probably corresponds to a change in the order of the phase transition¹¹ due to the temperature variation of the superconducting penetration depth $\lambda(T)$. For $T > T_d$, the phase boundary corresponds to a non-hysteretic second-order phase transition, whereas for $T < T_d$ the superconducting-normal phase transition is first order with a well defined hysteretic behaviour, as indicated by our R against

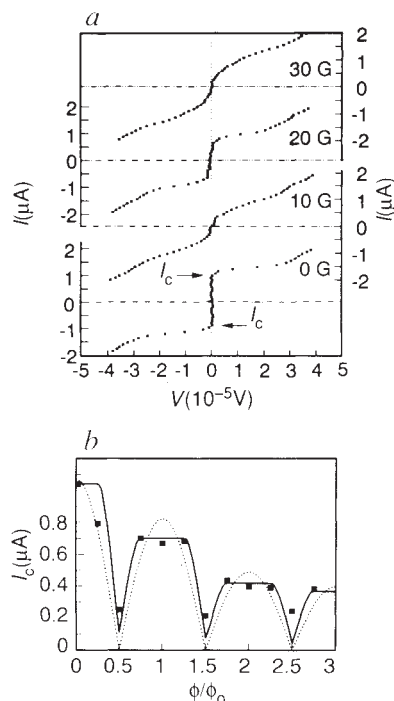


FIG. 3 *a*, The d.c. current-voltage characteristics of the $1 \mu\text{m} \times 1 \mu\text{m}$ square Al loop at $T = 1.30 \text{ K}$ for different applied magnetic fields. The left and right scales for the current I correspond to measurements in magnetic fields of 0, 20 G and 10, 30 G, respectively. The arrows illustrate the definition of the critical current I_c . *b*, The oscillations of the critical current I_c of the same Al loop as a function of normalized flux Φ/Φ_0 . The solid line corresponds to the calculations based on the theory of Fink *et al.*³ (see Fig. 1c), while the dashed line is the sinusoidal variation of I_c in a conventional SQUID with two Josephson junctions.

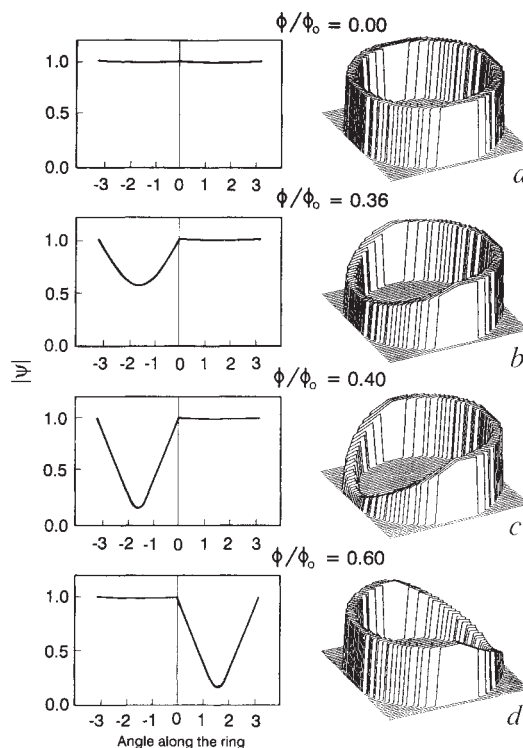


FIG. 4 *a-d*, Angular dependence of the modulus of the superconducting order parameter $|\psi|$ in a homogeneous mesoscopic loop for different normalized flux values Φ/Φ_0 (angle in radians). Three-dimensional plots of $|\psi|$ illustrate the formation of the area with strongly reduced $|\psi|$ value. This area plays the same role as artificially fabricated weak links in a conventional SQUID.

H measurements (not shown; for details see ref. 8). As a result, there is only a limited temperature interval ($1.26 \text{ K} \leq T \leq 1.32 \text{ K}$) in which the phase boundary oscillates weakly with Φ/Φ_0 , and where $R(H)$ corresponds to a second-order phase transition. For our Al loops the theoretical prediction by Fink *et al.* can only be checked within this limited temperature interval.

To test the validity of equation (5) we have measured the d.c. current-voltage ($I-V$) characteristics at $T = 1.3 \text{ K}$ for different values of the perpendicular magnetic field. The $I-V$ curves (Fig. 3a) are for the same loop for which the phase boundary in Fig. 2 was obtained. A strong dependence of the critical current amplitude on the field is observed. This dependence clearly shows the predicted oscillatory behaviour of I_c with normalized flux Φ/Φ_0 (Fig. 3b). As explained before, we can exclude the presence of extrinsic weak links as a possible source of these oscillations. Further confirmation comes from the fact that the characteristic product of the normal-state loop resistance and the critical Josephson current, $R_n I_c \approx 10^{-5} \text{ V}$, is at least two orders of magnitude smaller than typical values reported for loops with artificial Josephson junctions¹². We finally note that besides the expected oscillatory behaviour, the curve of I_c against Φ/Φ_0 shows the plateaus around integer flux quanta which are predicted by the nonlinear GL theory^{3,6} (solid line in Fig. 3b). On the other hand, our experimental results clearly deviate from the sinusoidal variation of $I_c(\Phi/\Phi_0)$ for a SQUID with weak links (dashed line in Fig. 3b).

As our numerical calculations show, the observed critical current oscillations are related to the presence of intrinsic 'weak links', the origin of which can be explained qualitatively as follows. The total current flowing through points (1) and (2) (Fig. 1b) is the sum of the expected circular shielding current I_s and the transport current I_t . The interplay between these currents produces a total current $I_t/2 + I_s$ at point (1) which is

larger than the total current $I_t/2 - I_s$ at point (2). A higher current amplitude decreases the modulus of the superconducting order parameter $|\psi|$, as illustrated in Fig. 4. For an integer Φ/Φ_0 , the shielding current $I_s = 0$ and we observe a symmetrical pattern for the order parameter in both branches of the loop (Fig. 4a). As the unbalance between the real flux Φ and the nearest integer flux $n\Phi_0$ increases, the shielding current increases and produces a suppression of $|\psi|$ in the branch where I_s and I_t are added and an enhancement of $|\psi|$ where I_s and I_t are subtracted (Fig. 4b). In the vicinity of the half-integer flux (Fig. 4c) this interference of I_s and I_t creates a 'self-made weak link' in one branch, which jumps into the other branch (Fig. 4d) as soon as $\Phi/\Phi_0 > 0.5$. The oscillatory behaviour of I_c in a mesoscopic superconducting loop may therefore be related to the creation of a region with a strongly reduced $|\psi|$ value ('weak-link'), switching between the two branches. This area plays the role of an artificial weak link in a conventional SQUID. Because of the phase coherence in the mesoscopic loop, the appearance of the 'weak-link' in one branch will simultaneously reduce the supercurrent in the second branch. $\square$

Received 23 October 1992; accepted 5 January 1993.

1. Josephson, B. D. *Phys. Lett.* **1**, 251–253 (1962).
2. Fink, H. J., López, A. & Maynard, R. *Phys. Rev.* **B26**, 5237–5240 (1982).
3. Fink, H. J., Grünfeld, V. & López, A. *Phys. Rev.* **B35**, 35–37 (1987).
4. de Gennes, P. G. *C. R. Acad. Sci. Ser. II* **292**, 279–282 (1981).
5. Alexander, S. *Phys. Rev.* **B27**, 1541–1557 (1983).
6. Fink, H. J., Loo, J. & Roberts, S. M. *Phys. Rev.* **B37**, 5050–5057 (1988).
7. Vloeberghs, H., Moshchalkov, V. V., Van Haesendonck, C., Jonckheere, R. & Bruynseraede, Y. *Phys. Rev. Lett.* **69**, 1268–1271 (1992).
8. Moshchalkov, V. V. *et al. Phys. Script.* **T45**, 226–229 (1992).
9. Little, W. A. & Parks, R. D. *Phys. Rev. Lett.* **9**, 9–13 (1962).
10. Tinkham, M. *Phys. Rev.* **129**, 2413–2422 (1963).
11. Arutyunyan, R. M. & Zharkov, G. F. *Zh. eksp. teor. Fiz.* **78**, 1530–1542 (1980); (Engl. transl.) *Sov. Phys. JETP* **51**, 768–774 (1980).
12. Muller, C. J., van Ruitenbeek, J. M. & de Jongh, L. J. *Phys. Rev. Lett.* **69**, 140–143 (1992).

ACKNOWLEDGEMENTS. We thank G. Neuttiens and H. Vloeberghs for their help with the phase boundary measurements, R. Jonckheere for sample preparation and L. Parijs for technical assistance. This research was supported by the Belgian Inter-University Institute for Nuclear Sciences (IIKW), the Inter-University Attraction Poles (IUAP), and the Concerted Action (GOA) programs. V.V.M. acknowledges financial support from the Research Council of the Katholieke Universiteit Leuven, M.D. and L.G. are Research Assistants of the Belgian Institute for the Encouragement of Scientific Research in Industry and Agriculture, and C.V.H. is Senior Research Associate of the Belgian National Fund for Scientific Research.

Evidence for the importance of bubbles in increasing air–sea gas flux

D. M. Farmer*, C. L. McNeil†‡ & B. D. Johnson‡

*Institute of Ocean Sciences, 9860 West Saanich Road, PO Box 6000, Sidney, British Columbia V8L 4B2, Canada

†University of Victoria, Victoria, British Columbia, Canada

‡Dalhousie University, Halifax, Nova Scotia, Canada

TWO models have been proposed to account for gas exchange between the atmosphere and the oceans: one involves direct transport of the gas through a surface boundary layer¹; the other also includes a substantial enhancement of the gas flux due to bubbles formed by breaking waves^{2,3}. In a long time-series of dissolved oxygen measurements, Wallace and Wirick⁴ observed sharply increased fluxes that seemed to be associated with wave activity. But the lack of vertical resolution meant that they could not rule out water advection and entrainment, rather than bubble-mediated air injection, as the cause of the increased flux. They were also unable to calculate transfer coefficients. Here we report simultaneous *in situ* observations from a vertical array of dissolved-gas sensors and a variety of other instruments during a single storm event. Our results confirm the importance of bubbles for the

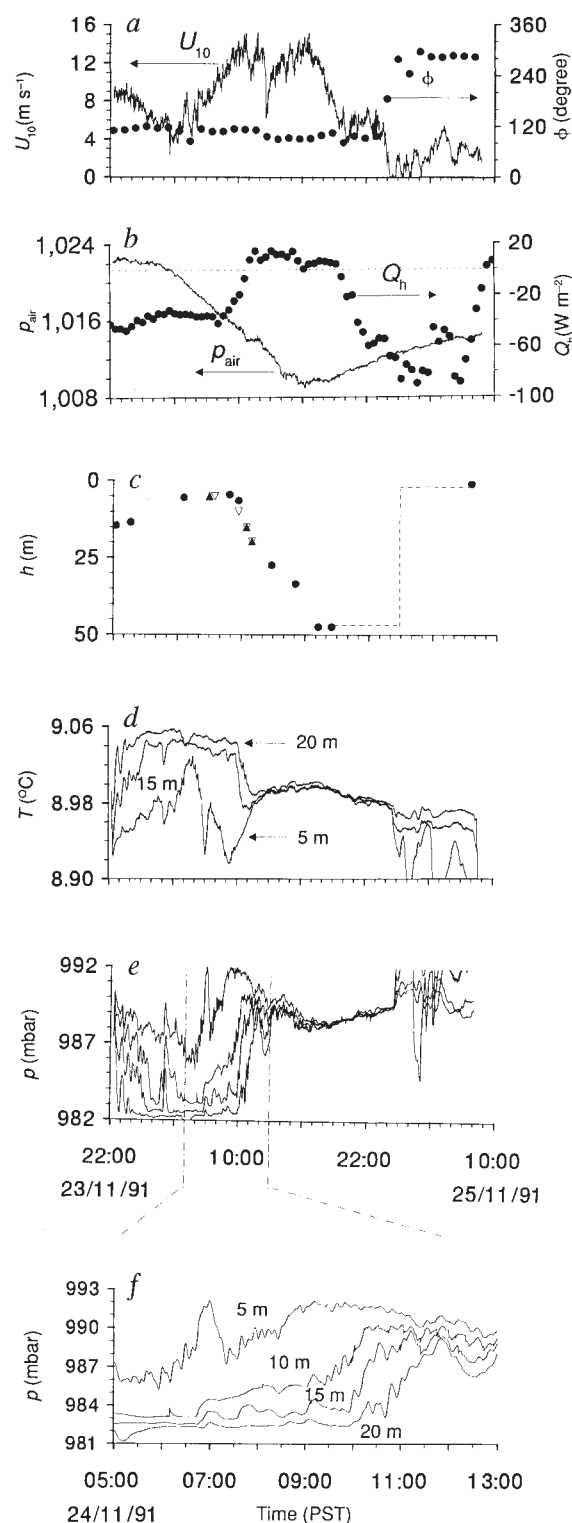


FIG. 1 Time series of observations in Georgia Strait (23–25 November 1991, 49° 46' N, 124° 45' W) showing (a) wind speed U_{10} (left) adjusted to 10 m and (right) wind direction ϕ ; (b) air pressure p_{air} (left) and air–sea heat flux Q_h (right); (c) mixed layer depth inferred from salinity/temperature profiles (●), moored thermistors (▲) and gas tension sensors (△); (d) water temperature at 5, 15 and 20 m; (e) gas tension at 5, 10, 15 and 20 m; and (f) an expansion of (e) showing progressive incorporation of sensors into the deepening mixed layer. The sensors are estimated¹⁷ to have an uncertainty in absolute measurement of 0.4%; corresponding d.c. offsets were applied to match gas tension within the well-mixed layer at 2000 h. A temperature sensitivity of 0.2 mbar over the observed temperature range was also found. Wind speed and mixed layer depth are estimated to have an uncertainty of $\pm 10\%$; water temperature is accurate to 0.02 °C.

gas-transfer process. They also imply that existing transfer coefficients underestimate the transfer of weakly soluble gases during periods of bubble penetration.

Our measurements were limited to a single storm, and thus complement the results of ref. 4 by providing a detailed analysis of dissolved gas in a surface mixing layer during a period of strong wind, with sufficient resolution to calculate exchange coefficients. Our observations were obtained during a storm in Georgia Strait, British Columbia, during November 1991. To minimize errors caused by advection past a fixed mooring, our sensors were suspended from a freely drifting float. The deployment consisted of a meteorological buoy, a multifrequency vertical sonar to detect bubbles⁵, an azimuthally scanning sonar (100 kHz, 3° s^{-1} sweep rate) for obtaining two-dimensional bubble cloud images, and a vertical array of four dissolved-gas sensors at depths of 5.1, 9.9, 15.1 and 19.7 m. Langmuir circulation was observed both with the azimuthally scanning sonar and with an acoustically tracked neutrally buoyant float. For an 8-h period during the height of the storm, a second nearby deployment included a breaking wave detector consisting of conductivity sensors at depths of 17, 32, 51 and 67 cm.

The gas tension sensor is an adaptation of an instrument described in ref. 6 and consists of a silastic membrane stretched over a rigid support containing a pressure and temperature sensor; it measures the sum of partial pressures of dissolved gases. Two days before the storm during calm conditions, dissolved oxygen concentration was measured by the Winkler titration method to 100 m simultaneously with gas tension at 20 m depth. The oxygen concentration was 19.9% undersaturated (with respect to 1 atm moist air⁷), which is typical of Georgia Strait in winter⁸. From these data the nitrogen concentration was estimated to be $5.3 \pm 0.8 \times 10^{-4} \text{ M kg}^{-1}$ (2.0% supersaturated).

The wind (Fig. 1a) was from the southeast, beginning at 0400 h, 24 November, and rising to 15 m s^{-1} . Before the storm, weak restratification occurred, appearing as a decrease in mixed layer depth (Fig. 1c). The depth then stabilized or increased only slightly during the period of rising wind, followed by more rapid increase at 0700 h. Bubble cloud penetration may initially have been inhibited by stratification (Fig. 2d). At 0530 h, coincident with the wind increase, the 5-m gas tension rose sharply until 0930 h, thereafter declining steadily until 1700 h. Deeper sensors initially showed little change until they were progressively incorporated in the mixed layer, first rising to the value at the 5-m sensor, and later falling in parallel with it. Evaluation of the heat and salt balance accurately reproduces the observed time dependence of salinity and temperature during the time of strong wind and rapid entrainment (0900–1800 h), thus justifying neglect of advection and use of a one-dimensional calculation for this limited period. Following the storm, fresher surface water appeared in the form of a sharp front, leading to rapid restratification of temperature, salinity and gas tension.

Figure 2 shows observations of the processes responsible for bubble generation and distribution. The significant wave height (Fig. 2b) reaches 1.7 m at 1500 h on 24 November with period 6.5 s. The fetch was 50 km. Wave-breaking frequencies (Fig. 2c) are comparable to previous observations at the same wind speed^{9,10}, when normalized by the dominant period. Bubble-cloud penetration depth (Fig. 2d) increases progressively with wind speed to 10 m. Figure 2e shows an azimuthally scanned sonar image with clear evidence of bubble clouds organized by Langmuir circulation into rows aligned with the wind. Neutrally buoyant float trajectories extended throughout the mixed layer with peak vertical velocities of 12 cm s^{-1} and r.m.s. values of $3\text{--}4 \text{ cm s}^{-1}$ (E. D'Asaro, personal communication). These observations confirm that wave-breaking, bubble injection and Langmuir circulation, which are the essential ingredients of bubble-enhanced gas transfer^{2,3,11}, all occurred during the storm.

The balance between the competing effects of gas transfer

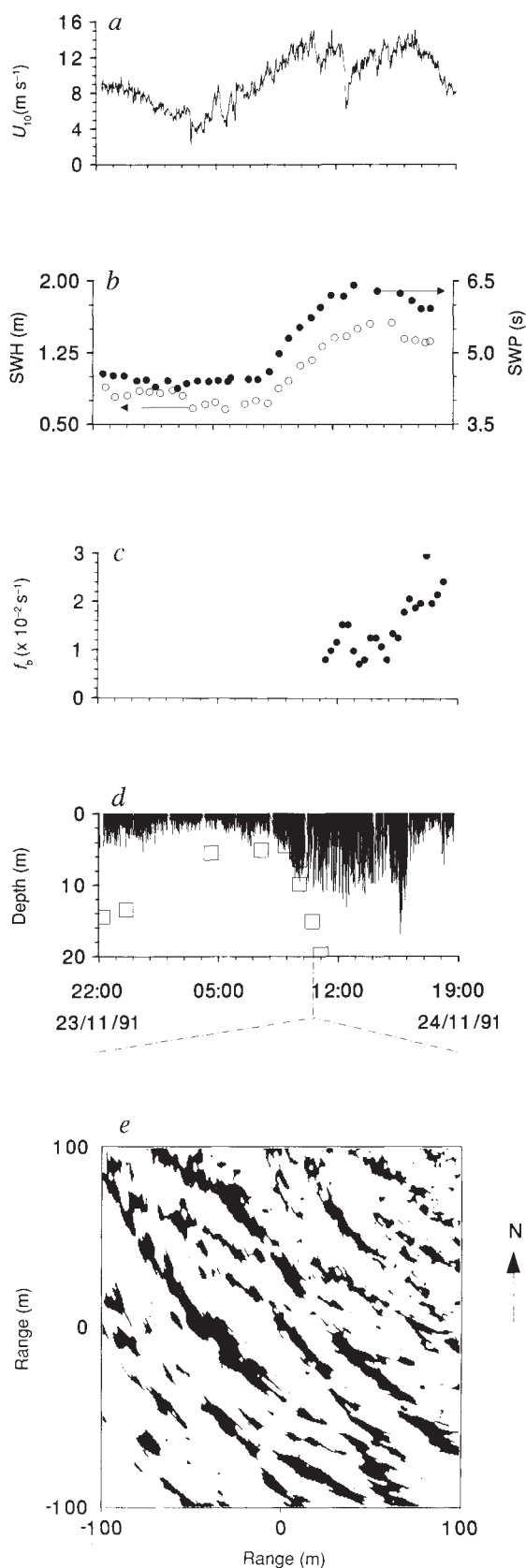


FIG. 2 Observations of wave conditions and bubble distribution. *a*, Wind speed U_{10} ; *b*, significant wave height (left) and significant wave period (right) measured with a vertical sonar; *c*, wave-breaking frequency, measured with conductivity sensors; *d*, bubble-cloud penetration depth determined from sonars, and mixed layer depth ($\square$); *e*, segment of scanning sonar image at 1030 h 24 November 1991, showing a two-dimensional view of bubble clouds organized by Langmuir circulation.

across the sea surface and entrainment, can be expressed by

$$h \frac{dc}{dt} = -Q - \frac{dh}{dt} (c - c_d) \quad (1)$$

where h is the mixed layer depth, c the surface layer dissolved gas concentration, Q the surface gas flux and c_d the entrained gas concentration. The direct transfer model¹ expresses the flux as a diffusive process

$$Q = -K_T(c_a - c) \quad (2)$$

where $K_T = K_T\{U_{10}, Sc\}$ is parameterized in terms of wind speed U_{10} and the Schmidt number Sc is defined as the ratio of kinematic viscosity to molecular diffusivity, c_a is the dissolved gas concentration in equilibrium with the atmosphere and depends on solubility $s = s\{\theta, S, p_{air}\}$ which is a function of temperature (θ), salinity (S) and atmospheric pressure (p_{air}). Woolf and Thorpe³ developed a model which includes both diffusive and bubble-induced effects

$$Q = -K_T(c_a(1 + \Delta_e) - c) \quad (3)$$

where $\Delta_e = \Delta_e\{U_{10}^2\}$ is the fractional supersaturation induced by bubbles for each gas constituent.

Insight into the resulting balance may be obtained from solutions of equations (1) and (2) or (3) for special cases of wind forcing and initial conditions. We take a constant initial stratification N and step wind-speed increase ($u_* = 0$ for $t < 0$, $u_* = U_*$ for $t \geq 0$); one-dimensional mixed layer models describe the resulting deepening. Niiler and Kraus¹² give $h = U_* N^{-1} (12mNt)^{1/3}$, where $m = 1.25$, before rotation becomes dominant. The solution to equations (1) and (3) for a single gas in this example is

$$c(\zeta) = c_s + i \left(\frac{\pi}{4} \right)^{1/2} (c_s - c_d) \zeta^{-1/2} \operatorname{erf}(i\zeta^{1/2}) e^{-\zeta} \quad (4)$$

where $c_s = c_a(1 + \Delta_e)$, $\zeta = (t/\tau)^{2/3}$ and $\tau = (32m/9)^{1/2} (U_*/K_T)^{3/2} N^{-1}$. Equation (4) (Fig. 3a) illustrates the rise or fall in concentration depending on initial conditions, asymptotically approaching the supersaturation level $\Delta_e(U_{10})$. In our observations, near-surface stratification and linearly increasing wind-speed reduce initial entrainment. A more representative result for this case is obtained by solving equations (1) and (3) for linear wind increase and quadratic deepening, with $c_w/c_s = 0.8$. The modelled gas concentration first rises, but then drops (Fig. 3b) as entrainment dominates in qualitative agreement with the observations (Fig. 1e).

Numerical integration of equation (1) with (2) or (3) for observed wind forcing (Fig. 1a) and observed mixed-layer depth (Fig. 1c) provides a basis for direct comparison. The gas flux equations are solved separately for oxygen and nitrogen, allowing for changes in solubility⁷ due to water temperature, salinity and atmospheric pressure, and then combined with water vapour pressure $p_{water}(\theta)$ and argon using Dalton's law together with Henry's law to calculate gas tension. Windspeed and mixed layer depths are represented by functional approximations. Salinity at 15 m increased from 29.0 to 29.5 units during the storm. In the absence of observations above 5 m, calculations begin when the 5-m sensor is incorporated in the surface layer. By assuming that nitrogen concentration is independent of depth and equal to 5.3×10^{-4} M kg⁻¹ (from prior measurement), and 100% argon saturation, the gas tension at each sensor before the storm (0500 h, 24 November) allows calculation of the initial oxygen profile to 20 m. Once the mixed layer has penetrated beneath 20 m, we have only the deeper observations of dissolved oxygen obtained two days before the storm, so the calculations are less certain.

Both the model of Woolf and Thorpe³, and the direct transfer model of ref. 1 (W&T, DT in Fig. 3c) underpredict the observations. There is great scatter in published values of K_T (refs 3, 13, 14). We therefore fit the model of ref. 3 to the observations by multiplying K_T by a factor α as a function of time (Fig. 3d).

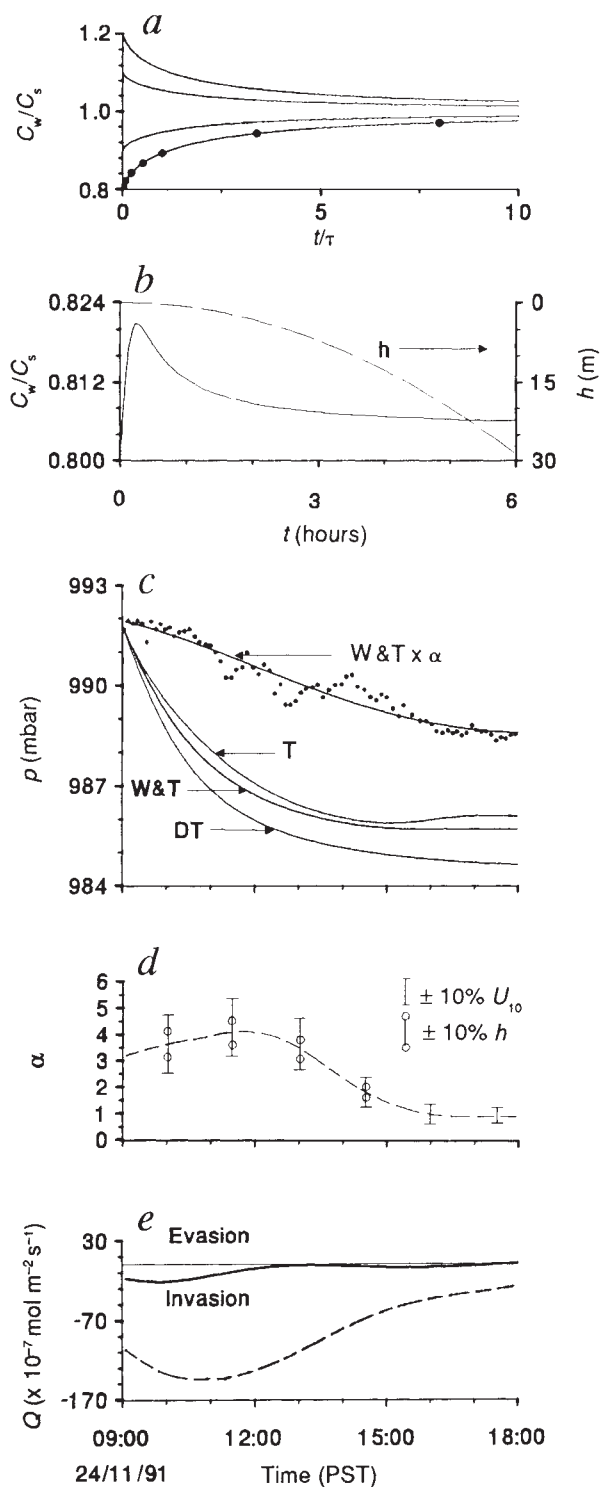


FIG. 3 Model gas concentrations and fluxes. *a*, The solution of equations (1) and (3) for a step function increase in wind speed, showing changes in gas concentration relative to saturation level $\Delta_e(U_{10})$; points (●) show verification of the numerical integration scheme by direct evaluation of the analytical solution (4) using tabulated values of the error function¹⁸. *b*, As for *a* but with linearly increasing wind speed and quadratic mixed layer deepening, showing initial rise in gas concentration followed by decrease. *c*, Observations (●) and model predictions of surface layer gas tension: models include direct transfer^{1,3} (DT), Woolf and Thorpe³ (W&T), and Thorpe¹⁹ (T). The smooth curve through the gas tension measurements is a quartic least-squares fit, which is used in *d* to define the corresponding enhancement α of the gas transfer coefficient as a function of time. Thus the quadratic fit in *c* corresponds to W&T $\times \alpha$ where $\alpha(t)$ is given in *d*. Error bars are shown for $\pm 10\%$ uncertainty in windspeed and mixed layer depth. *e*, Inferred fluxes of N_2 (solid curve) and O_2 (dashed), from W&T $\times [\alpha(t)]$.

The enhancement is greatest ($\alpha \approx 4$) when the bubble clouds penetrate most deeply (Fig. 2d). Calculated surface fluxes of oxygen (Fig. 3e) are comparable to previous observations⁴. The nitrogen flux is less than that of oxygen, because nitrogen is close to saturation and oxygen significantly undersaturated. Both gases are invasive during strong winds, but there is slight nitrogen evasion at low winds.

There are several reasons for expecting departures from previously published values of K_T . Apart from limitations of estimates based only on gas evasion, the influence of fetch and averaging period can introduce significant bias¹⁵. Many previous field studies exclude entrainment which, as shown here, can dominate the time dependence of gas concentration, leading to

underestimates, especially in stratified coastal environments. Application of photographic measurements¹⁶ of bubble size distributions, which miss the smaller bubbles, could have led to an underestimate of bubble-induced gas transfer in ref. 3. Acoustical measurements⁵ suggest a spectral peak in bubble radius of less than 20 μm , possibly increasing modelled bubble-induced transfer. Detailed calculations require analysis of bubble distributions, Langmuir circulation and related phenomena (Fig. 2). Our initial results, indicating considerable enhancement of K_T during bubble penetration, provide strong support for the role of bubble injection in the transfer of weakly soluble gas, and emphasize the relevance of near-surface oceanography to the estimation of transfer coefficients. $\square$

Received 10 August 1992; accepted 12 January 1993.

1. Liss, P. S. *Phil. Trans. R. Soc. A* **325**, 93–103 (1988).
2. Thorpe, S. A. *Phil. Trans. R. Soc. A* **304**, 155–210 (1982).
3. Woolf, D. K. & Thorpe, S. A. *J. mar. Res.* **49**, 435–466 (1991).
4. Wallace, D. W. R. & Wirrick, C. D. *Nature* **356**, 694–696 (1992).
5. Vagle, S. & Farmer, D. M. *J. Atmos. Ocean Tech.* **9**(5), 630–644 (1992).
6. Anderson, M. L. & Johnson, B. D. *J. geophys. Res.* **97**, 17899–17904 (1992).
7. Weiss, R. F. *Deep-Sea Res.* **17**, 721–735 (1970).
8. Crean, P. B. & Ages, A. B. *Oceanographic Records from Twelve Cruises in the Strait of Georgia and Juan de Fuca Strait, Vol. 1* (Department of Energy, Mines and Resources, Marine Sciences Branch, Victoria BC, 1968).
9. Thorpe, S. A. & Humphries, P. N. *Nature* **283**, 463–465 (1980).
10. Lamarre, E. & Melville, W. K. *IEEE J. Ocean Engng* **17**, 204–215 (1992).
11. Thorpe, S. A. *J. Fluid Mech.* **142**, 151–170 (1984).
12. Niller, P. P. & Kraus, E. B. *One-Dimensional Models of the Upper Ocean, Modelling and Prediction of the Upper Layers of the Ocean* 143–172 (Pergamon, Oxford, 1977).

13. Liss, P. S. & Merlivat, L. in *The Role of Air-Sea Exchange in Geochemical Cycling* (ed. Buat-Menard, P.) 113–127 (Reidel, Dordrecht, 1986).
14. Watson, A. J., Upstill-Goddard, R. C. & Liss, P. S. *Nature* **349**, 145–147 (1991).
15. Wanninkhof, R. *J. geophys. Res.* **97**, 7373–7382 (1992).
16. Johnson, B. D. & Cooke, R. C. *J. geophys. Res.* **84**, 3761–3766 (1979).
17. Johnson, B. D. *Calibration Procedures and Evaluation of Performance for the Model 1 GTD Field Instrument*, Tech. Rep. (Dept. of Oceanography, Dalhousie University NS, 1992).
18. Abramowitz, M. & Stegun, A. *Handbook of Mathematical Functions* (Dover, New York, 1970).
19. Thorpe, S. A. *Ann. Geophys.* **2**, 53–56 (1984).

ACKNOWLEDGEMENTS. This work was supported through the Canadian Joint Global Ocean Flux Study, the U.S. Office of Naval Research and the Canadian Panel on Energy Research and Development. C.L.M. was supported by the Royal Society of Edinburgh. We thank A. D. Booth for discussions, D. Slauenwhite for assistance with calibrations, J. Gemmrich for analysis of wave-breaking statistics and M. Trevorrow for assistance in acoustical analysis.

A previously unrecognized late-glacial cold event in eastern North America

André J. Levesque, Francis E. Mayle, Ian R. Walker* & Les C. Cwynar

Department of Biology, University of New Brunswick, Bag Service Number 45111, Fredericton, New Brunswick, Canada E3B 6E1
* Biology Department, Okanagan University College, North Kelowna Campus, Kelowna, British Columbia, Canada V1Y 4X8

THE transition between the last glacial period and the present interglacial was marked by pronounced and abrupt changes in climate^{1,2}, one of which, the Younger Dryas event, resulted in a return to almost full-glacial conditions. Here we present evidence for a short but intense cold period in eastern North America, pre-dating the Younger Dryas, revealed by analysing the organic, pollen and midge content of lake sediments at very high resolution. We name this event the Killarney Oscillation and it lasted from 11,160 to 10,910 radiocarbon years BP. Because this corresponds to the Gerzensee Oscillation in Switzerland and can be correlated with other short-term events reported from across the North Atlantic seaboard^{3–14}, we conclude that a single event affected all of the North Atlantic region. This pre-Younger Dryas cold period is similar in nature and geographic extent to the major cooling of the Younger Dryas¹⁵, raising the possibility that these two events are causally related.

The study area comprises two of the three Canadian Maritime provinces (43°–48°N, 60°–68°W), New Brunswick and Nova Scotia. Deglaciation of the region may have begun as early as 14,500 yr BP and was completed by 11,000 yr BP, although some evidence suggests renewed glaciation associated with the Younger Dryas cooling event¹⁶, which has been radiocarbon-dated by accelerator mass spectrometry (AMS) as 10,770–10,000 yr BP¹⁷. At 11 of 12 lake sites examined for the present study, there is a clear decrease in the organic content of sediment, as measured by loss-on-ignition (LOI) at 550 °C, before onset

of the Younger Dryas (Fig. 1) and coinciding with a lighter coloration of the sediment. The increased minerogenic component of the sediment is interpreted as evidence for decreased productivity and/or increased inwash or solifluction as a result of reduced vegetation cover, both processes reflecting a response to climate cooling. The lower and upper boundaries for the Killarney Oscillation (KO) in all diagrams are defined using each site's LOI diagram: the point where the percentage LOI begins to drop is taken as the lower KO boundary and the point where it returns to that value, or levels off, as the upper KO boundary.

For six of the sites for which we have detailed LOI diagrams, we have also completed high-resolution pollen analysis (0.5–1 cm intervals) across the interval bearing the KO (Fig. 2). Before the KO, southern New Brunswick and central Nova Scotia were forested by either spruce (*Picea*) or poplar (*Populus*) woodland, whereas central New Brunswick and southern and northern Nova Scotia were dominated by shrub-tundra. At Joe Lake, percentages of poplar and willow (*Salix*) decline during the KO as sedges (Cyperaceae), total herbs and dwarf birch (*Betula glandulosa*) increase (>300 dwarf birch fruits and catkin scales and no tree birch macrofossils have been recovered from late-glacial sediment from six sites in Maritime Canada, so we attribute all late-glacial birch pollen to dwarf birch). The disappearance of poplar and increase in herbs and shrubs suggest an opening up of the vegetation in response to cooling. The Killarney Lake diagram shows a peak in the abundance of poplar (over 50%) before the onset of the KO. The extremely rapid decline in poplar and contemporaneous increase in dwarf birch, willow, Cupressaceae (probably *Juniperus communis*, judging from needles recovered from late-glacial sediment from Little Lake and Lac à Magie¹⁷), sedge and total herbs suggest a reversion from poplar woodland to shrub-tundra. Before the KO at Pine Ridge Pond, percentages of spruce, pine (*Pinus*), dwarf birch and poplar were roughly the same (20–25%). The decline in poplar and pine, and pause in the population growth of spruce, coincident with a rise in dwarf birch and Cupressaceae, indicate an increase in shrubs at the expense of trees in response to cooling. At Stillman Pond, high percentages of dwarf birch, Cupressaceae and sweet gale (*Myrica*) and the

presence of poplar indicate that essentially shrub-tundra, with a minor tree component, was established before the start of the KO. The decline of these taxa and concomitant increase in sedge, grass (Poaceae) and total herbs indicate a reversion from shrub to herb-tundra. The same interpretation is given to the pollen changes at Chase Pond, where sedge, grass and total herbs increase at the expense of dwarf birch. The Main-à-Dieu Pond diagram shows only a consistent increase in sedge before the onset of the Younger Dryas. As none of the other taxa decline significantly, we interpret these changes to indicate a slight opening of the landscape in response to cooling. Although the vegetation changes associated with the KO at all sites are consistent with the interpretation of a climate cooling, the intensity of the response varies from site to site. The smallest changes occur at Joe Lake and Main-à-Dieu Pond, and can be explained by the relative insensitivity of the vegetation that existed there before the onset of the KO (mainly shrub/herb-tundra), or perhaps a less intense cooling as the relative changes in LOI are smallest at those sites.

Analysis of midge remains from Killarney Lake (Fig. 3) reveals peak abundances of cold-stenothermous Chironomidae such as *Heterotrissocladius*, *Paracladius* and *Protonypus* during the KO and Younger Dryas, whereas taxa characteristic of warm littoral habitats, such as temperate-littoral Chironomini and

Pentaneurini, were least abundant during these intervals. When a recent model for the reconstruction of summer surface-water temperature¹⁸ is applied to these chironomid assemblages, the resulting curve clearly shows temperature minima associated with both the KO and Younger Dryas (Fig. 3), with lake summer surface-water temperature dropping by 4.5 °C during the KO. Although a substantial deepening of the lake might produce similar chironomid changes, it is unlikely that the late-glacial water levels of Killarney Lake differed appreciably from today's; cores taken near the shore contain complete late-glacial records, and the maximum water level is set by a bedrock sill, above which excess water drains.

Lower and upper boundary age estimates for the KO (Table 1) were calculated using AMS ¹⁴C dates from Joe Lake, Mayflower Lake, Splan Pond, Killarney Lake and Stillman Pond. In all cases, the material dated consisted primarily of unidentifiable twig fragments, occasionally supplemented by spruce needles or other terrestrial plant macrofossils. Nearly all dates represent a vertical increment of only 1 cm, and not exceeding 2 cm. They place the mean start of the KO at ~11,160 yr BP and its end at ~10,910 yr BP, indicating that the KO lasted ~250 years. Boundary estimates from Chase Pond imply that the KO lasted for ~1,000 years and started more than 1,000 years earlier than at any other of our dated sites. This seems unlikely given

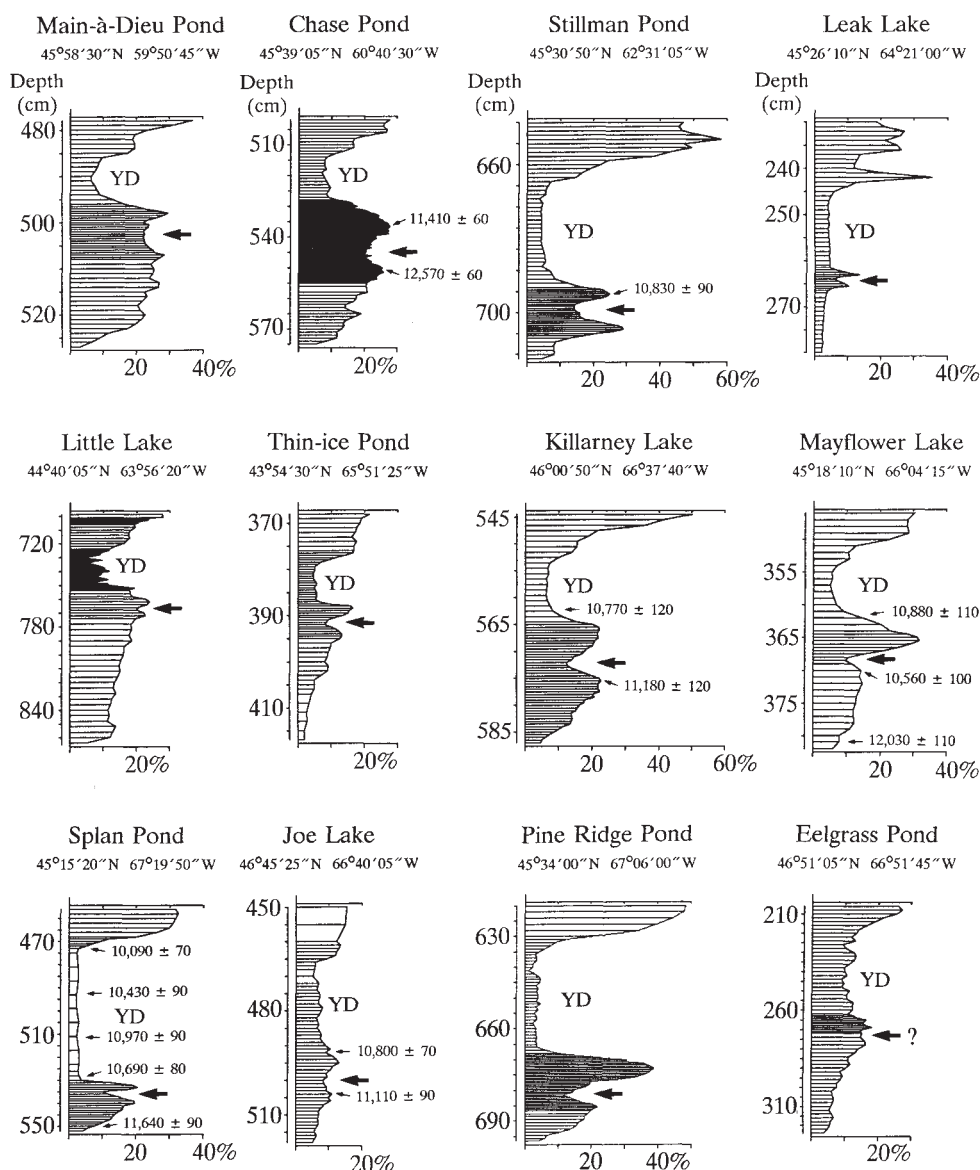


FIG. 1 Loss-on-ignition (LOI) curves for sediment from lakes in Maritime Canada. The large arrows point to the decline in LOI before the onset of the Younger Dryas (YD) cooling event. Also shown are AMS ¹⁴C dates used to estimate boundary ages for the KO. Each horizontal line of the LOI curves represents one sample.

TABLE 1 Age estimates for Killarney Oscillation

Site	Sample depth (cm)	Sample description	Weight (mg)	IsoTrace laboratory no.	Age (yr BP)	Lower boundary age estimate (yr BP)	Upper boundary age estimate (yr BP)
Joe Lake	491.5–492.5	twig fragments	57	TO-2302	10,800 ± 70	11,120	10,920
	503.5	twig fragments	14	TO-2303	11,110 ± 90		
Killarney Lake	562.5	twig fragments	89	TO-3002	10,770 ± 120	11,200	10,930
	575	twig fragments	18	TO-3001	11,180 ± 120		
Splan Pond	471.5–473.5	bract, seed, twig and leaf fragments	28	TO-1933	10,090 ± 70	11,190	10,950
	492.5–493.5	twig fragments	52	TO-1932	10,430 ± 90		
	511.5–512.5	twig fragments	150	TO-1931	10,970 ± 90		
	529.5–531.5	twig fragments, spruce needles	34	TO-1930	10,690 ± 80		
	551–553	twig and conifer needle fragments	24	TO-1928	11,640 ± 90		
Mayflower Lake	361.5–362.5	twig fragments, spruce needles	96	TO-1948	10,880 ± 110	11,130	10,930
	369.5–370.5	twig fragments	14	TO-2351	10,560 ± 100		
	380.5–381.5	twig fragments	36	TO-1947	12,030 ± 110		
Stillman Pond	690.5–691.5	twig fragments	11	TO-3055	10,830 ± 90		10,830
Chase Pond	541.5–543.5	twig fragments	26	TO-2326	11,410 ± 60	12,570*	11,630*
	555.5–556.5	twig fragments	17	TO-2325	12,570 ± 60		

Lower and upper boundary age estimates of the Killarney Oscillation for sites in Maritime Canada determined using AMS ^{14}C dates from each site (AMS ^{14}C dates from Joe Lake, Splan Pond, Mayflower Lake and Chase Pond have been reported previously¹⁷). Boundary ages for Joe Lake, Killarney Lake, and Chase Pond were calculated by linear interpolation or extrapolation between or from AMS ^{14}C dates. Boundary ages for Splan Pond and Mayflower Lake were calculated by linear interpolation between the basal AMS ^{14}C date and the age for the lower Younger Dryas boundary¹⁷. A single AMS ^{14}C date provides the upper KO boundary estimate for Stillman Pond. Age estimates with an asterisk are considered anomalous. The AMS ^{14}C dates are the average of two machine-ready targets and have been corrected for natural and sputtering fractionation to a base of $\delta^{13}\text{C} = -25\text{‰}$. The sample ages are quoted as uncalibrated conventional radiocarbon dates in years before present (yr BP), using the Libby ^{14}C meanlife of 8,033 years. The errors represent 68.3% confidence intervals.

the proximity of sites and stratigraphic consistency of this event, and because Chase Pond AMS ^{14}C dates are probably anomalous¹⁷, we have rejected these estimates.

Oxygen isotope studies on late-glacial sediment from numerous lake sites in Switzerland revealed a consistent drop in the $^{18}\text{O}/^{16}\text{O}$ ratio just before the onset of the Younger Dryas^{3–5}. This has been interpreted as a widespread cooling event, the Gerzensee Oscillation. The end of the Gerzensee Oscillation is preceded by the deposition of the Laacher See tephra, a well established time marker of 11,000 yr BP¹⁹. The

end of the Gerzensee Oscillation therefore coincides with the end of the Killarney Oscillation which has been dated at 10,910 yr BP.

The Gerzensee Oscillation has been correlated^{13,14} with a variety of pre-Younger Dryas cooling events in the North Atlantic. High-resolution analysis of deep-sea cores with high sedimentation rates from the North Atlantic¹³ and the Norwegian Trench¹⁴ revealed a peak in the percentage abundance of the polar species of foraminifer *N. pachyderma* (sinistral coiling) before the start of the Younger Dryas. Stratigraphically

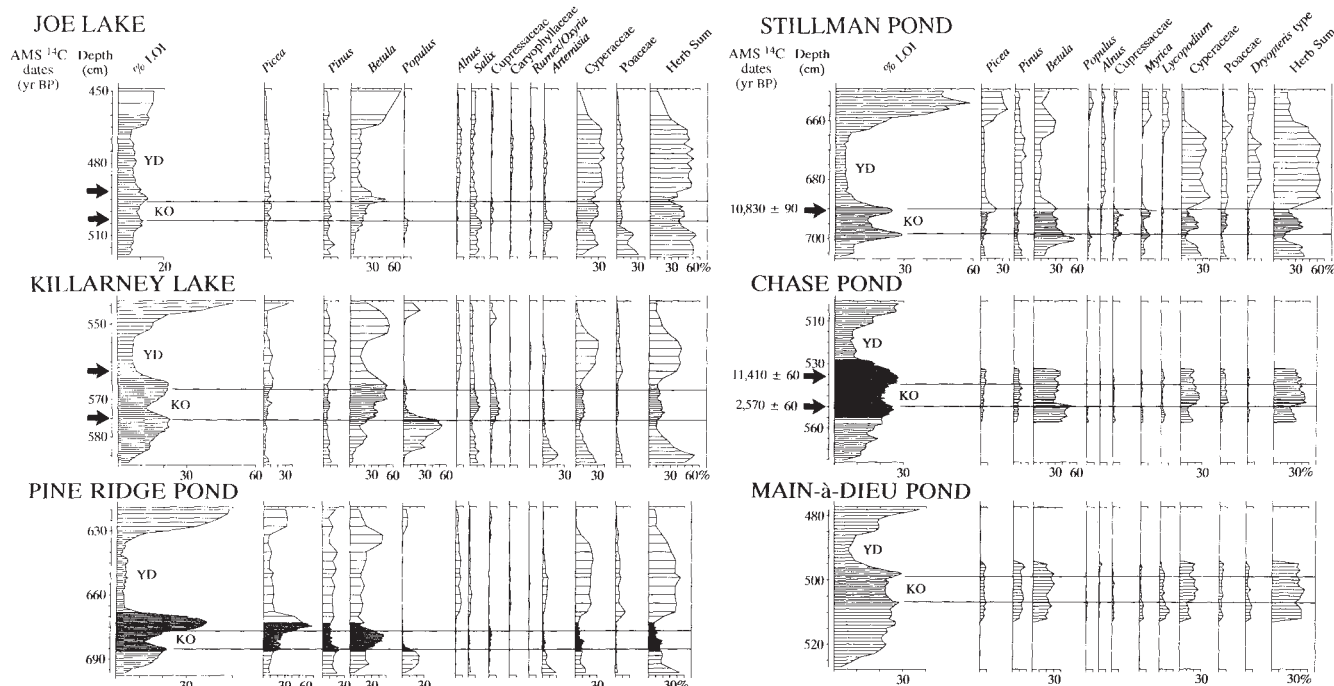
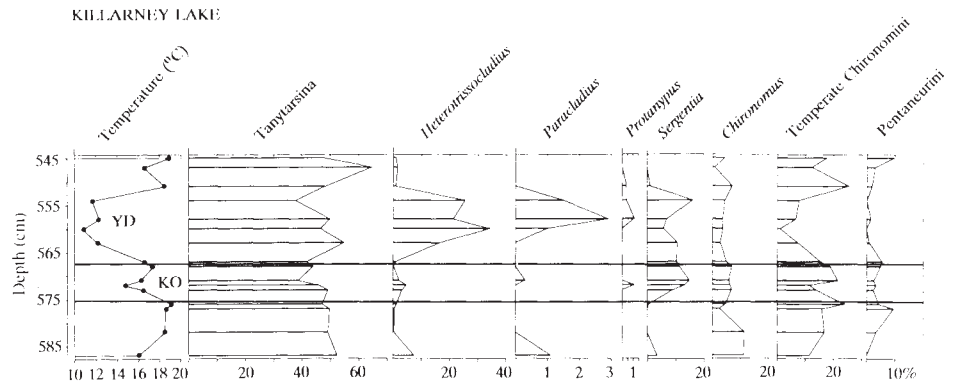


FIG. 2 Pollen percentage diagrams from six sites in Maritime Canada. The thick continuous horizontal lines indicate the boundaries of the KO based

on the LOI curves, as discussed in the text. YD, Younger Dryas. Each horizontal line of the LOI and pollen curves represents one sample.

FIG. 3 Late-glacial changes in summer surface-water temperature and relative abundance of chironomid taxa from Killarney Lake. The apparent root-mean-square error of the temperature estimates is 1.51 °C (ref. 18). Boundaries of the KO are based on the LOI curves and are represented by the thick, continuous horizontal lines. YD, Younger Dryas.



similar fluctuations have been recorded in oxygen isotope profiles from the Summit, Dye 3, Camp Century, and Renland Greenland ice cores⁶⁻⁸, and pollen profiles from northwest Europe⁹⁻¹². Given that the Gerzensee Oscillation corresponds with the Killarney Oscillation and that similar but geographically intermediate events have been correlated with the Gerzensee Oscillation^{13,14}, the inference is that a single, short but intense cooling event affected oceanic, atmospheric and terrestrial systems across the entire North Atlantic region.

Because the late-glacial climate oscillation proposed here was short-lived (~250 yr) and is represented by just 7.0 cm of sediment on average in Maritime Canada, it may have gone undetected in North America and elsewhere by conventional sampling at 5 or 10 cm intervals. The occurrence of small-scale (KO type) and medium-scale (Younger Dryas) events that can be correlated across the North Atlantic points to the close coupling of late-glacial climates across the region. □

Received 17 August 1992; accepted 5 January 1993.

- Berger, W. H. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **89**, 219-237 (1990).
- Broecker, W. S. *Quat. Res.* **38**, 135-138 (1992).
- Eicher, U. *Mitt. Naturforsch. Ges. Bern* **37**, 65-80 (1980).
- Eicher, U. & Siegenthaler, U. *Boreas* **5**, 109-117 (1976).
- Lotter, A., Ammann, B. & Sturm, M. *Clim. Dyn.* **6**, 233-239 (1992).
- Siegenthaler, U., Eicher, U., Oeschger, H. & Dansgaard, W. *Ann. Glaciol.* **5**, 149-152 (1984).
- Oeschger, H. et al. *Geophys. Monogr.* **29** (eds Hansen, J. E. & Takahashi, T.) 299-306 (AGU, Washington, DC, 1984).
- Johnsen, S. J. et al. *Nature* **359**, 311-313 (1992).
- Riezebos, P. A. & Slotboom, R. T. *Boreas* **13**, 347-353 (1984).
- Iversen, J. *Danm. geol. Unders.* **7**, 345-445 (1973).
- Paus, A. *Boreas* **17**, 113-139 (1988).
- Paus, A. *J. Quat. Sci.* **4**, 223-242 (1989).

- Broecker, W. S. et al. *Paleoceanography* **3**, 1-19 (1988).
- Lehman, S. J. & Keigwin, L. D. *Nature* **356**, 757-762 (1992).
- Mott, R. J., Grant, D. R., Stea, R. & Occhietti, S. *Nature* **323**, 247-250 (1986).
- Stea, R. R. & Mott, R. J. *Boreas* **18**, 169-187 (1989).
- Mayle, F. E., Levesque, A. J. & Cwynar, L. C. *Quat. Res.* (in the press).
- Wilson, S. E., Walker, I. R., Mott, R. J. & Smol, J. P. *Clim. Dyn.* (in the press).
- Bogaard, P. & Schmincke, H.-U. *Geol. Soc. Am. Bull.* **96**, 1554-1571 (1985).

ACKNOWLEDGEMENTS. We thank J. J. Lowe, H. H. Birks, H. J. B. Birks and R. W. Spear for comments on an earlier version of this paper, T. J. Keenan and L. Doner for help with fieldwork and S. L. McLean for technical assistance. This work was supported by a NSERC Canada research grant to L.C.C. The AMS ¹⁴C dates were determined by the IsoTrace laboratory at the University of Toronto, a facility supported by NSERC.

Deformation of a weak subducted slab and variation of seismicity with depth

Winston C. Tao & Richard J. O'Connell

Harvard University Department of Earth and Planetary Sciences,
20 Oxford Street, Cambridge, Massachusetts 02138, USA

OCEANIC lithosphere is assumed to possess platelike properties whether it is lying at the Earth's surface or descending deep into the mantle; yet the geometries of subducting slabs indicate significant deformation as they descend through the mantle¹⁻⁴, suggesting that lithospheric plates might be so weakened during subduction as to act not as a rigid solid, but as a viscous fluid. Numerical and laboratory experiments have shown that fluid 'slabs' can indeed take on realistic profiles⁵⁻⁷; however, it has not been clear whether a fluid or 'weak' model of slab dynamics can account for deep earthquakes, which are usually ascribed to the deformation of a rigid or 'strong' slab. Here we show, using numerical simulations of slab evolution, that a weak slab model is in fact consistent with seismic observations. Assuming that earthquakes occur at a rate proportional to deformation rate, we reproduce the observed variation with depth of seismicity rate and focal mechanisms, and the cessation of seismicity at 670 km depth. Provided that sinking material encounters resistance at depth (here modelled as a viscosity jump at 670 km), the pattern of seismicity can be explained by any mechanism for deep earthquakes in which the rate of seismicity is proportional to strain rate in the slab.

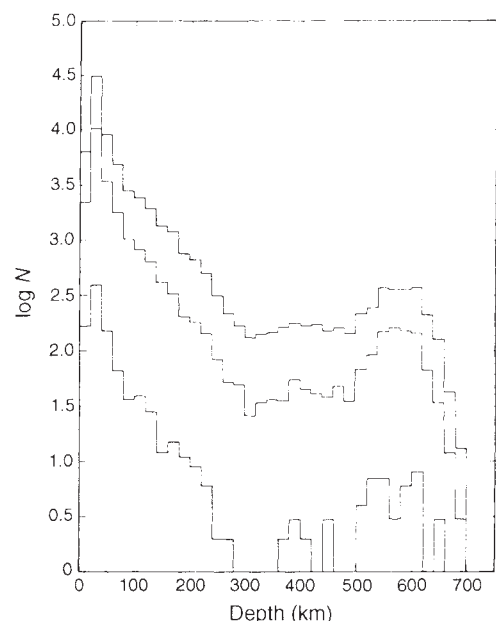


FIG. 1 Logarithm of number of earthquakes worldwide against depth (1964-1980), after ref. 8. Curves are for cutoff magnitudes of (left to right) $m_b \geq 4$, $m_b \geq 5$ and $m_b \geq 6$. Seismicity rates decrease exponentially with depth from the Earth's surface to a minimum at 300-500 km; in most subduction zones, focal mechanisms in this region of decay indicate down-dip extension. In the deep slab, seismicity strongly increases in a region of down-dip compression, then abruptly ceases in the lower mantle.

Our experiments represent the endmember case of the 'weak slab' model, which we define as one in which all lithospheric strength is lost when a plate enters a trench, and a subducted 'slab' is distinguished from surrounding material solely by its greater density. We use these experiments to determine how slabs will deform if their shapes are entirely controlled by patterns of mantle flow (that is, if slab rigidity is negligible compared with mantle-flow-induced stresses). From these deformations we infer seismicity, assuming that seismicity rates vary with the slab's thickness and with its instantaneous rate of down-dip strain.

Our assumption differs from that of earlier numerical studies, in which seismicity rates were correlated with the stress field in a high-viscosity fluid slab⁸. As the origin of deep seismicity is uncertain, the true relation between slab stress, strain rate and seismicity is unclear; we assume a strain-rate dependence here because it is the simplest relation that accommodates a wide variety of deep earthquake models. Proposed earthquake mechanisms include thermally induced shear instabilities⁹, whose activity will vary with strain rate; plastic instabilities¹⁰ and localized polymorphic phase transitions^{11,12}, which should show a stress dependence; and the dewatering of hydrated minerals entrained by the slab¹³, a process requiring only some minimum stress to act as a trigger. The thermal and plastic instability models suggest that seismicity can occur throughout the body of a slab, and a strain-rate dependence would then apply either directly (thermal instability) or through the dependence of stress on strain rate in a fluid slab (plastic instability). The remaining mechanisms imply that earthquakes occur in a narrow, chemically distinct seismogenic zone, perhaps associated with the elastic upper portion of the lithosphere. Stresses within the seismogenic zone might then differ from those in the rest of the fluid slab; however, if the deformation of this zone is controlled by the slab's overall distortions, the pattern of stress within the seismogenic region should parallel the overall strain rate.

The seismic observations of interest here are summarized in Fig. 1. Earthquakes within subducted slabs vary with depth in both focal mechanism orientation and seismicity rate. Deep focal mechanisms (at depths greater than 300–400 km) almost always indicate down-dip compression, whereas down-dip extension is more typical of intermediate (70–300 km) depths^{8,14–16}. Seismicity rates decay exponentially with depth in the uppermost mantle, reach a minimum between depths of 300 and 500 km, then increase between 500 and 700 km (Fig. 1). Below ~700 km, seismicity abruptly ceases; this is often cited as an argument against slab penetration into the lower mantle, although it has also been proposed that slabs penetrate a seismically¹⁷.

Our experimental model was originally developed for the study of ablative subduction; for details, see ref. 5. We begin with a two-dimensional cartesian box with periodic sidewalls, the bottom of the box representing the core–mantle boundary and the top representing the top of the viscous lithosphere (Fig. 2). A laterally homogeneous newtonian viscous fluid fills the box, with viscosity transitions at 90 km and 670 km delineating

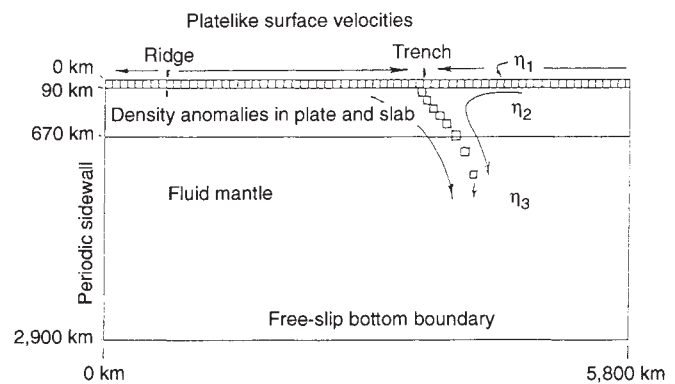


FIG. 2 Geometry and main features of the experimental model, from ref. 5. A fluid mantle is overlain by a gravitationally unstable fluid lithosphere, whose surface moves in a platelike fashion. The density anomaly of the lithosphere is modelled as a series of discrete density anomalies, each of which freely advects with mantle and lithospheric flow. When anomalies sink, they form a fluid 'slab' of the same viscosity as surrounding material. Plates are driven by slab-induced flow, and the trench and ridge migrate over time.

the viscous lithosphere, upper mantle and lower mantle. The fluid that makes up the viscous lithosphere is denser than that of the mantle, and the sinking of this fluid leads to the formation of strengthless 'slabs'. Atop the viscous lithosphere, a platelike velocity boundary condition simulates the effect of an elastic lithosphere separated into two periodic plates by a migrating trench and ridge^{18–20}. Each experiment is begun by perturbing an initial amount of material downwards at the trench. As subducted material sinks into the mantle, it generates a flow field that drives the plates together and draws still more material downwards to form a growing slab. Plate motions are thus slab-driven: although there is no mechanical connection between slab and plate, surface and interior motions are coupled by the mantle flow, which acts as a viscous form of slab pull. Slab density is constant as material sinks, and the evolving shape of the fluid slab is controlled by mantle flow.

Although strengthless, the slabs thus generated adopt realistic profiles. We plot two slab profiles in Fig. 3, one generated by subduction in a uniform mantle, the other produced using a 10-fold mantle viscosity increase at 670 km. Streamlines in both cases show that material accelerates as it draws away from the overlying plates and is entrained by the flow surrounding the slab; this initial acceleration leads to an extension and thinning of subducted material. In a uniform mantle, the region of extension passes through and beyond 670 km and the deep slab shows no down-dip compression. In a viscosity-stratified mantle, extension occurs only at shallow and intermediate depths; at greater depths, compression arises as flow is impeded by the increased viscosity below 670 km. The latter deformation pattern qualitatively resembles that seen in other fluid studies^{5–7}, and is consistent with that observed in most subduction zones.

FIG. 3 Fluid slabs generated in (a) a uniform mantle and (b) a mantle whose viscosity increases by a factor of 10 at 670 km. Only a section of the subducted plate is shown; ridges lie out of sight, and the overriding plate is omitted for clarity. Rectangles of constant area show the positions of density anomalies; the rectangle widths vary inversely with anomaly spacing, and hence indicate the cumulative down-dip strain within the slab. Together with the instantaneous streamlines, these rectangles show how extensional deformation may occur as the slab sinks away from the surface, and that compression of the deep slab can occur if mantle viscosity increases with depth.

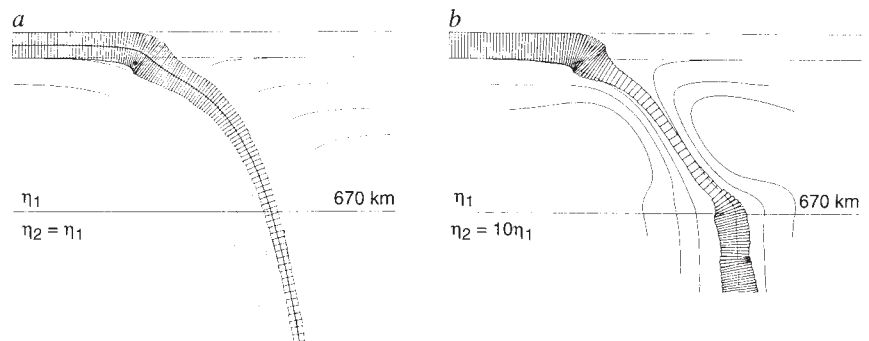
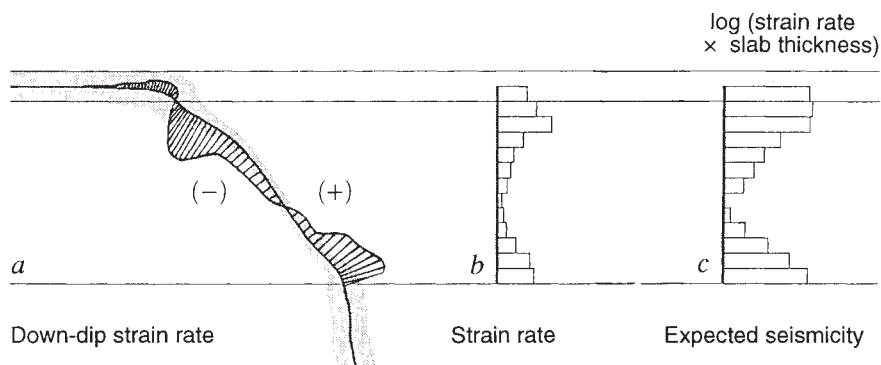


FIG. 4 Strain rates and expected rates of seismicity in a fluid slab. *a*, Strain rate within slab of Fig. 3*b* (mantle viscosity increases by factor of 10 at 670 km). Strain rate is plotted using centre of slab as axis; curve to the right of the axis indicates compression (+), to the left indicates extension (-). Minimal deformation occurs mid-slab at the crossover from extension to compression, and the most vigorous strain rates occur as the slab draws away from the surface and as it nears the viscosity step at 670 km. The areas of maximal strain rate coincide with the regions of greatest slab thickness. Below 670 km, down-dip strain rates drop precipitously. *b*, Integral of strain rates (integrated down-dip) at 50-km depth intervals. *c*, Logarithm of expected seismicity, calculated by taking the integral of the down-dip strain rate multiplied by slab thickness. In the weak-slab model, the number of earthquakes peaks where strain rates and slab thickness are greatest. The resulting curve resembles that of Fig. 1: there is an exponential decay with depth from the surface to



the midslab, corresponding to a region of downdip extension, followed by a compressional resurgence at depth. Below 670 km, the abrupt drop in strain rate may allow the slab to sink aseismically.

From the deformation history of our experimental slabs, we calculate the expected variation of seismicity with depth (Fig. 4). We assume that all parts of the lithosphere are firmly bound together and hence deform at equal rates. Whether earthquakes are distributed throughout a slab or are concentrated in a seismogenic layer, the number of earthquakes within a given depth interval will be proportional to the amount of slab material present and to the rate of down-dip strain.

In Fig. 4*a* we plot the instantaneous strain rate for the slab in Fig. 3*b*. Extensional and compressional peaks are seen towards the top and bottom of the slab, with a decrease in strain rate in the cross-over region in between and an abrupt drop in strain rate beneath 670 km. Integrating this strain rate along the slab over successive depth intervals gives Fig. 4*b*. To obtain the expected seismicity, we must take into account the thickness of the slab (and hence of the seismogenic region) at different depths. We therefore multiply the integrated strain rate by slab thickness to obtain the expected seismicity distribution in Fig. 4*c*.

Given the extreme simplicity of the experimental model, the calculated seismicity pattern (Fig. 4*c*) is remarkably similar to that observed (Fig. 1). In both, the rate of seismicity decreases exponentially with depth in the upper slab, achieves a minimum at intermediate depths and increases towards 670 km in a region of down-dip compression. Moreover, our results suggest a simple mechanism for aseismic slab penetration into the lower mantle: if mantle viscosity increases moderately at 670 km, slab strain rates abruptly decrease below the viscosity interface (Fig. 4*a*), yielding an effective cessation of seismicity even though the slab continues to descend (Fig. 4*c*).

The fluid-slab model thus allows for the formation of realistic slab profiles, accounts for the depth variation in seismic focal mechanisms observed in most subduction zones, yields the observed pattern of seismicity rate variation with depth, and suggests a simple explanation for how a slab might penetrate aseismically into the lower mantle.

The model does not explain how slabs may undergo continuous compression at all depths; such a deformation pattern, although not common, has been observed beneath Tonga-Kermadec¹⁴ and may perhaps arise from complications of mantle flow created by complex plate and slab geometries, a possibility our simple two-dimensional experiments cannot address. Moreover, the mechanisms behind deep earthquakes are uncertain, and observed seismicity patterns might reflect depth variations in the pressure, temperature or slab composition required for deep seismicity. In this respect, however, the weak-slab model makes it easier to account for deep earthquakes; rather than requiring that a given earthquake mechanism explain the observed variations in seismicity, we see that any mechanism that correlates seismicity with strain rate may be consistent with observation. The ability to allow simple explanations of major

phenomena is the most intriguing aspect of the weak-slab model, and suggests that slabs should be viewed not as strong bodies thrust beneath a trench, but instead as weak fluid sheets flowing down into the Earth's mantle. □

Received 13 August 1992; accepted 11 January 1993.

1. Bevis, M. *Nature* **323**, 52-53 (1986).
2. Fischer, K. & Jordan, T. H. *J. geophys. Res.* **96**, 14429-14444 (1991).
3. Giardini, D. & Woodhouse, J. H. *Nature* **307**, 505-509 (1984).
4. Giardini, D. & Woodhouse, J. H. *Nature* **319**, 551-555 (1986).
5. Tao, W. & O'Connell, R. J. *J. geophys. Res.* **97**, 8877-8904 (1992).
6. Gurnis, M. & Hager, B. H. *Nature* **335**, 317-321 (1988).
7. Kincaid, C. & Olson, P. *J. geophys. Res.* **92**, 13832-13840 (1987).
8. Vassiliou, M. S., Hager, B. H. & Raefsky, A. A. *J. Geodyn.* **1**, 11-28 (1984).
9. Ogawa, M. *J. geophys. Res.* **92**, 13801-13810 (1987).
10. Hobbs, B. E. & Ord, A. *J. geophys. Res.* **93**, 10521-10540 (1988).
11. Green II, H. W., Young, T. E., Walker, D. & Scholz, C. H. *Nature* **348**, 720-722 (1990).
12. Kirby, S. H. *J. geophys. Res.* **92**, 13789-13800 (1987).
13. Meade, C. & Jeanloz, R. *Science* **252**, 68-72 (1991).
14. Apperson, K. D. & Frohlich, C. *J. geophys. Res.* **92**, 13821-13831 (1987).
15. Isacks, B. & Molnar, P. *Nature* **223**, 1121-1124 (1969).
16. Isacks, B. & Molnar, P. *Rev. Geophys. Space Phys.* **9**, 103-174 (1971).
17. Creager, K. C. & Jordan, T. H. *J. geophys. Res.* **91**, 3573-3589 (1986).
18. Gable, C. W., O'Connell, R. J. & Travis, B. J. *J. geophys. Res.* **96**, 8391-8405 (1991).
19. Hager, B. H. & O'Connell, R. J. *J. geophys. Res.* **86**, 4843-4867 (1981).
20. Ricard, Y. & Vigny, C. *J. geophys. Res.* **94**, 17543-17559 (1981).

ACKNOWLEDGEMENTS. Supported by the NSF and NASA. We thank the Department of Geology and Geophysics at Princeton University for the use of computer facilities.

Aerodynamics and the evolution of long tails in birds

Andrew Balmford*, Adrian L. R. Thomas & Ian L. Jones*

Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK

Two problems limit the interpretation of recent experiments¹⁻³ supporting Darwin's⁴ suggestion that female choice for ornate males may account for the evolution of long tails in birds. First, in some species tail elongation may have been favoured by natural rather than sexual selection. Second, it is unclear how female preferences for elaborate males have evolved, because current tests of competing models are often inconclusive⁵⁻⁷. We have integrated aerodynamics theory with comparative data on sexual dimorphism in tail length to evaluate the flight costs of different forms of tail elongation. We report here that long tails with shallow forks are aerodynamically optimal, exhibit correspondingly low sexual dimorphism and may therefore have evolved under natural selection. Other long-tail types impair flight and show greater sexual

*Present addresses: Institute of Zoology, Zoological Society of London, Regent's Park, London NW1 4RY, UK (A.B.); Department of Zoology, University of British Columbia, 6270 University Boulevard, Vancouver, British Columbia V6T 2A9, Canada (I.L.J.).

dimorphism, but variation in their initial evolutionary cost suggests differences in how female preferences for them may have evolved.

Although evidence is accumulating that in a number of birds females prefer long-tailed mates¹⁻³, the role of natural selection in tail elaboration has been largely ignored. Evidence that long tails impose viability costs is very limited⁸⁻¹⁰, yet without this it remains possible that tail elongation has sometimes been driven by natural selection, perhaps because it improves aerodynamic efficiency, manoeuvrability or stability¹¹. Moreover, even where it has been demonstrated that long tails are costly but are preferred by choosy females, resolving how such preferences have evolved is difficult, particularly if females gain no direct benefits from their choice of mating partners⁵. For example, evidence that preferred traits are costly, that they reveal heritable variation in male quality and that discriminating females thereby produce high viability offspring, would be consistent with 'good genes' models of sexual selection¹². But such results could also be accommodated entirely by Fisherian models⁵⁻⁷, in which initially arbitrary male traits become increasingly elaborate and costly as a result of coevolution with female preferences for them¹³. An unexplored way to distinguish these competing models may be to switch from examining the contemporary costs of male traits and the benefits of female preferences for them, and to consider instead the costliness of traits during the initial stages of their evolution. Modern Fisherian models make no firm predictions because runaway evolution can occur whether preferred traits are initially costly, beneficial or neutral with respect to natural selection¹³. In contrast, good genes models require that if their evolution is gradual, preferred characters must incur substantial costs even in their simplest form to evolve as reliable indicators of male viability.

We have used a combination of aerodynamic models and comparative data from more than 600 museum skins to examine the costs of tail elongation in birds. Our analysis is based on a new lifting surface model¹⁴ which applies the classical aerodynamic solution of the forces acting on a low aspect ratio delta wing^{15,16} to assess the lift and drag associated with a bird's tail. Tails generate lift in proportion to the square of their maximum continuous span. Any area distal to the point of maximum width contributes only drag, which is proportional to total tail area¹⁴. The aerodynamically optimal tail is therefore triangular when spread, and forked when closed. We examined the aerodynamic implications of different types of tail elongation by modelling progressive elaboration from a short, simple tail in which all feathers are of equal length (see Fig. 1).

Long, graduated tails, in which all but the outermost feathers are elongated, generate the same lift as normal tails (because maximum span is set by the outermost feathers and so does not change) but substantially more drag, as elongation adds considerably to tail area. The aerodynamic cost of the tail (measured here as drag/lift at a constant angle of attack, though other indices give qualitatively similar results) therefore rises sharply with increased length (Fig. 1). In pintails only the central tail feathers are elongated, so that although lift again fails to increase with length, tail area and thus drag rise only slowly. The marginal aerodynamic costs of tail elongation (that is costs per unit length) are therefore relatively slight for pintails. Lastly, if tails become forked (when closed) through elongation of their outer feathers, lift initially increases more rapidly than drag (as the spread tail tends toward the optimum triangular shape), so that aerodynamic costs are lower for long shallow forks than for simple tails¹⁴. Costs remain essentially unchanged if these forks continue to elongate isometrically, but if elongation instead proceeds by further deepening of the fork through lengthening its outer but not its central feathers, these outer feathers eventually project beyond the point of maximum continuous width even when spread. Drag (but not lift) then increases with length, so that aerodynamic costs rise, albeit at the same low rate as for pintails. Note that use of a more detailed analytical method such as curved lifting line theory¹⁷⁻¹⁹ generates qualitatively

TABLE 1 Comparisons of tail length in closely related pairs of long-tailed species

(a) Forked versus other tail types		Relative tail length		
Species pair	Tail shape	Male	Female	Male:Female
<i>Sterna paradisaea</i>	Fork	1.14	1.04	1.10
<i>Stercorarius longicaudus</i>	Pin	1.37	1.34	1.02
<i>Eupetomena macroura</i>	Fork	1.84	1.73	1.06
<i>Phaethornis superciliosus</i>	Pin	1.52	1.44	1.06
<i>Merops hirundineus</i>	Fork	1.24	1.24	1.00
<i>Merops superciliosus</i>	Pin	1.16	1.11	1.05
<i>Coracias abyssinica</i>	Fork	2.01	1.81	1.1†
<i>Uratelornis chimaera</i> *	Graduated	2.35	1.77	1.33
<i>Gubernates yetapa</i>	Fork	2.63	2.34	1.12
<i>Alecturus risorius</i>	Pin	2.55	1.59	1.60
<i>Phibalura flavirostris</i>	Fork	1.30	1.17	1.11
<i>Illicura militaris</i>	Pin	1.06	0.69	1.54
<i>Motacilla clara</i>	Fork	1.22	1.30	0.94
<i>Euplectes macrourus</i>	Graduated	1.30	0.62	2.10
<i>Enicurus immaculatus</i>	Fork	1.54	1.54	1.00
<i>Copsychus malabaricus</i>	Graduated	2.10	1.59	1.32
<i>Dicrurus macrocercus</i>	Fork	1.58	1.47	1.07
<i>Elminia longicauda</i>	Graduated	1.93	1.71	1.13

(b) Shallower versus deeper forks		Relative tail length		
Species pair	Male fork depth	Male	Female	Male:female
<i>Chelictinia ricourii</i>	1.93	1.57	1.48	1.06
<i>Elanoides forficatus</i>	2.53	1.56	1.45	1.08
<i>Sterna melanogaster</i>	2.23	1.23	1.30	0.95
<i>Sterna paradisaea</i>	2.54	1.14	1.04	1.10
<i>Hemiprocne comata</i>	1.92	1.22	1.27	0.96
<i>Hemiprocne longipennis</i>	2.20	1.21	1.01	1.20
<i>Polyonyxus caroli</i> †	1.37	1.32	1.14	1.16
<i>Sappho sparganura</i>	2.95	1.99	1.15	1.73
<i>Coracias caudata</i>	1.64	1.34	1.26	1.06
<i>Coracias abyssinica</i>	1.95	2.01	1.81	1.11
<i>Muscipira vetula</i>	1.45	1.49	1.44	1.03
<i>Gubernates yetapa</i>	3.62	2.63	2.34	1.13
<i>Hirundo rustica</i>	2.27	1.30	1.21	1.07
<i>Hirundo smithii</i>	2.44	1.19	0.79	1.51
<i>Dicrurus caerulescens</i>	1.30	1.38	1.42	0.97
<i>Dicrurus macrocercus</i>	1.55	1.58	1.47	1.07

All pairs are from the same family or superfamily^{21,22} and are roughly matched for size and relative tail length of males, but have otherwise been selected at random. Body length and the lengths of the longest and shortest tail feathers were measured on 10 male and 10 female skins in breeding plumage (except* = only 5 males and 6 females; † = only 8 females). To account for overall size differences between specimens, relative tail length (mean length of left and right longest tail feathers/body length) was calculated, and the ratio of the mean relative tail length of males to that of females then used as an index of sexual dimorphism in tail length. For forked tails, depth is the ratio of the mean lengths of their longest and shortest feathers. In these and all other analyses, we have excluded all species whose long tails may have mechanical functions other than flight (such as acting as a brace when climbing).

similar conclusions to those of our simple general lifting surface approach¹⁴.

This analysis yields a number of testable results. It suggests that, depending on their depth (longest:shortest tail feather length) and how widely they are spread during use, long forked tails may be aerodynamically beneficial and so (unlike other long tails) may have evolved through natural rather than sexual selection. Given that sexual selection in birds usually acts more strongly on males than females⁴, we predict that the reduced role of sexual selection in the evolution of long forked tails should mean that these generally exhibit lower sexual dimorphism in length than do long graduated tails or pintails. We tested this by measuring museum skins to compare tail dimorphism (male:female tail length, standardized to body length) in pairs of closely related, long-tailed species, which differed in the shape of their long tails but in few other potentially confounding variables. The results of this pairwise comparative approach²⁰

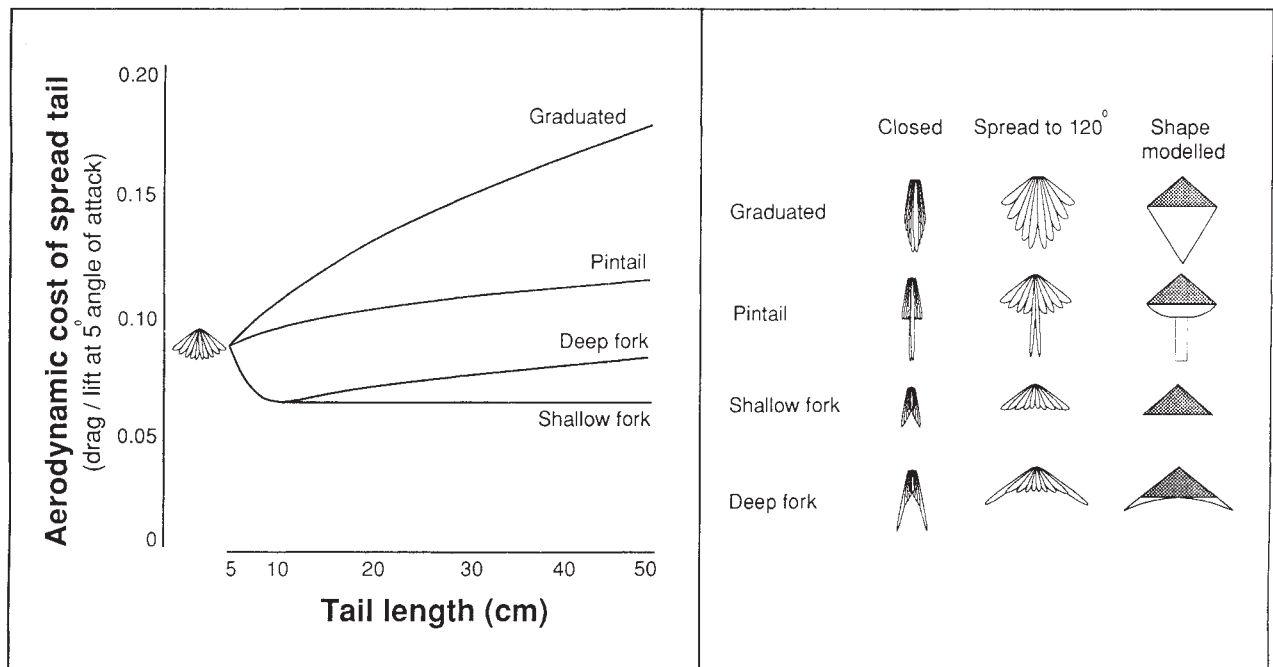


FIG. 1 Aerodynamic costs associated with the progressive elongation of different types of long tails from a simple tail with all feathers 5 cm long. Lift and drag have been calculated for tails spread to an apical angle of 120°. The tail shapes considered are shown in the right-hand panel, together with the corresponding geometric polygons used in the model; stippling indicates what part of each tail generates lift. The marginal costs of elongation of pintails and deep forks will depend only on the width of their streamers, but will always be substantially lower than those of graduated tails, in which nearly all tail feathers are elongated. Tails are assumed to

act as stiff flat plates with continuous surfaces¹⁴. Aerodynamic costs are expressed as the ratio of drag/lift at 5° angle of attack and 5 m s⁻¹ freestream velocity. The airflow over the tail is assumed to be a steady uniform flow comprising the vector sum of the bird's velocity and the induced velocity generated by the wings. Flight is assumed to be either normal forward flapping flight^{23,24}, or gliding²⁵. Conclusions are nevertheless robust across a broad array of flight speeds and flow conditions and are unaffected by the precise details of the tail shapes considered.

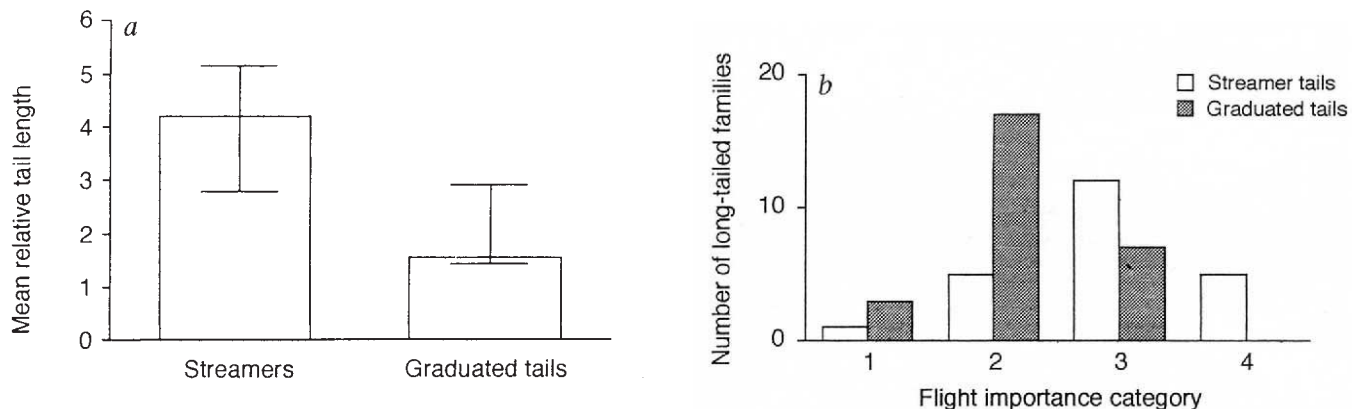


FIG. 2 *a*, Mean relative tail lengths of the species with the longest streamer tail (pintail or deep fork) and graduated tail in each of the seven families that contain both long tail types (Psittacidae, Caprimulgidae, Trochilidae, Tyrannidae, Monarchidae, Ploceidae and Paradisaeidae). The plot gives median values and interquartile ranges across all families. Relative tail lengths are based on three specimens per species of the sex with the longest tail (males in 13 species; females in one hummingbird showing reversed sexual dimorphism). *b*, The number of long-tailed families with streamers or graduated tails in relation to the importance of flight in the family. Each family was categorized by its commonest long tail type and its median score for flight importance, assessed by three ornithologists blind to the tail length data. Flight importance categories are: (1) almost entirely terrestrial; (2) largely terrestrial and feed while clambering, but fly between food patches; (3) often rest but rely on aerial manoeuvrability to obtain food; (4) largely aerial and feed on wing. *c*, The number of long-tailed families with streamers or graduated tails compared with whether or not they migrate. Tendency to migrate was again assessed by three independent ornithologists.

confirm our prediction (Table 1a): across nine species pairs, forked tails do indeed show less sexual dimorphism in length than do other long tails (Wilcoxon signed ranks test: $z_{\text{corr}} = 2.10$, 1-tailed $P = 0.017$).

Our model also suggests that as long forked tails deepen, it becomes more likely that they are aerodynamically costly (compared with shallow forked tails; see Fig. 1), and therefore increasingly probable that sexual rather than natural selection explains their evolution. If this is so, then among species with long forked tails, sexual dimorphism in tail length should increase with fork depth. Pairwise comparisons of closely related species support this prediction (Table 1b). Even where differences in fork depth are slight, these are consistently associated with greater sexual dimorphism in the tails of more deeply forked species (Wilcoxon test across $N = 8$ family pairs: $z_{\text{corr}} = 2.52$, 1-tailed $P = 0.006$).

Among those elongated tails that are apparently sexually selected, the marginal costs of graduated tails greatly exceed those of streamer-shaped tails (that is pintails and deep forks; Fig. 1). This generates several predictions. Within families, the lower costs per unit length of streamer tails should mean that they have evolved to a greater maximum length than more expensive, graduated tails. This is indeed the case (Fig. 2a): in all seven families with both long-tail types, the maximum length of streamer tails (relative to body length) exceeds that of graduated tails (Wilcoxon test: $z = 2.37$, 1-tailed $P = 0.009$). The last two predictions deal with differences between bird families in their reliance on flight, and therefore in the likely fitness consequences of increased aerodynamic costs. Comparing long-tailed families, as birds spend more time flying and rely more on aerial agility when feeding, we should expect graduated tails to become less common than streamers. Similarly, costlier graduated tails should be less prevalent in families that migrate than in other families. Both of these predictions are confirmed across 50 long-tailed families, categorized according to 'flight importance' and tendency to migrate by three independent ornithologists (Fig. 2b, c). Families that spend a substantial proportion of time flying and/or feed aerially (flight importance categories 3 and 4) are far less likely to exhibit graduated tails than are more terrestrial families (categories 1 and 2; $\chi^2 = 11.46$, d.f. = 1, $P < 0.001$; Fig. 2b). Likewise, where long-tailed birds migrate, they are more likely to possess streamer-shaped tails than are non-migrant birds ($\chi^2 = 8.86$, d.f. = 1, $P < 0.01$; Fig. 2c).

Each of the predictions of our aerodynamic analysis is therefore supported by the comparative data, enabling us to make two suggestions about the selective pressures driving tail elongation in birds. First, the theoretical and morphological evidence indicates that while most other types of long tails have probably evolved through sexual selection, natural selection for aerodynamic efficiency may be a sufficient explanation for the evolution of many elongated forked tails. Second, the marginal costs associated with pintails and deep forks are far lower than those of graduated tails. Clearly, even short-tail streamers might represent considerable handicaps in families where the fitness consequences of impaired flight are extremely high. But in other families where flight is less important and both streamers and graduated tails have evolved, our analysis suggests that during their initial evolution, streamers would have been relatively poor indicators of male quality. Thus Fisherian models could account for their initial elaboration, but good genes models probably could not. In contrast, graduated tails would have constituted uncheatable handicaps even in their simplest form, and so are more likely to have evolved as reliable indicators of the viability of potential mates. □

6. Balmford, A. & Read, A. F. *Trends Ecol. Evol.* **6**, 274–276 (1991).
7. Jones, I. L. *Auk* **109**, 197–199 (1992).
8. Møller, A. P. *Nature* **339**, 132–135 (1989).
9. Evans, M. R. & Hatchwell, B. J. *Behav. Ecol. Sociobiol.* **29**, 421–427 (1992).
10. Evans, M. R. & Thomas, A. L. R. *Anim. Behav.* **43**, 337–347 (1992).
11. Maynard Smith, J. *Evolution* **6**, 127–129 (1952).
12. Heisler, L. in *Sexual Selection: Testing the Alternatives* (eds Bradbury, J. W. & Andersson, M.) 96–118 (Wiley, Chichester, 1987).
13. Lande, R. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3721–3725 (1981).
14. Thomas, A. L. R. *Phil. Trans. R. Soc.* (in the press).
15. Jones, R. T. National Advisory Committee for Aeronautics, Technical Report **835**, (1946).
16. Jones, R. T. *Wing Theory* (Princeton Univ. Press, Princeton, 1990).
17. Guermond, J. L. *J. Fluid Mech.* **211**, 497–513 (1990).
18. Cheng, H. K. & Murillo, L. E. *J. Fluid Mech.* **143**, 327–350 (1984).
19. van Dam, C. P. *Nature* **325**, 435–437 (1987).
20. Felsenstein, J. *Am. Nat.* **125**, 1–15 (1985).
21. Sibley, C. G., Alquist, J. E. & Monroe, B. L. *Jr Auk* **105**, 409–423 (1988).
22. Howard, R. & Moore, A. *A Complete Checklist of the Birds of the World* 2nd edn (Academic, London, 1991).
23. Rayner, J. M. V., Jones, G. & Thomas, A. L. R. *Nature* **321**, 162–164 (1986).
24. Spedding, G. R. *J. exp. Biol.* **127**, 59–78 (1987).
25. Spedding, G. R. *J. exp. Biol.* **127**, 45–57 (1987).

ACKNOWLEDGEMENTS. A.B. and A.L.R.T. contributed equally to this work. We thank the staff of the Natural History Museum, Tring, the University Museum of Zoology, Cambridge, and the American Museum of Natural History in New York for access to their collections; M. de L. Brooke, W. Duckworth and M. Evans for scoring flight behaviour; S. Blakeman and A. Willmott for help with measurements; and M. de L. Brooke, N. Davies, C. Ellington, D. Harper, P. Harvey, K. McComb, A. Read and A. Willmott for helpful comments. A.P.B. was supported by Girton College, Cambridge, A.L.R.T. by SERC, and I.L.J. by NSERC, Canada.

Landmark stability is a prerequisite for spatial but not discrimination learning

R. Biegler & R. G. M. Morris*

Laboratory for Neuroscience, Department of Pharmacology, University of Edinburgh Medical School, 1 George Square, Edinburgh EH8 9JZ, UK

NEURONS sensitive to both place and direction from distinct regions of the hippocampal formation^{1,2}, allometric relationships between spatial learning and hippocampal structure^{3,4} and pronounced impairments in spatial learning after lesions in this area^{5–8}, indicate that the hippocampal formation subserves allocentric spatial learning^{9,10}. To learn more about the process of spatial representation, we have developed a task that provides independent control of both landmark and directional cues. On the basis of physiological¹¹ and behavioural¹² work, this task also makes it possible to investigate the relevance of associative learning principles, such as predictability¹³, to the spatial domain. We report here that although rats learn to discriminate between landmarks on the basis of their proximity to a reliably predicted food reward, they will only learn to use them to represent its location if they maintain stable locations within a geometric frame of reference.

We trained 14 male Lister-hooded rats to find food reward in a large directionally polarized arena (Fig. 1a). Two landmarks (L+, L−) were arranged to predict the spatial location of hidden food nearby that was either accessible (F+) or inaccessible (F−). The single parameter that we manipulated was whether the landmarks and feeders occupied fixed locations across trials (group fixed) or were moved randomly from trial to trial (group varied). Importantly, when the landmarks were moved between locations, the feeders were also moved such that the spatial relationship between L+ and F+, and between L− and F−, was always maintained; that is, if F+ was 40 cm to the 'south' of L+ for a given rat, it remained so on all trials.

Spatial theories predict that, because only group fixed has the potential of forming a stable geometric representation of the landmarks (that is, 'map'), only they will learn to search at the F+ location. Conversely, application of the associative learning principle of predictability to the spatial domain suggests that both groups should be able to learn to search at F+, but

Received 21 October; accepted 31 December 1992.

1. Andersson, M. *Nature* **299**, 818–820 (1982).
2. Møller, A. P. *Nature* **332**, 640–642 (1988).
3. Andersson, M. *Anim. Behav.* **43**, 379–388 (1992).
4. Darwin, C. *The Descent of Man, and Selection in Relation to Sex* (Murray, London, 1871).
5. Kirkpatrick, M. & Ryan, M. J. *Nature* **350**, 33–38 (1991).

* To whom correspondence should be addressed.

the association between L+ and F+ should be greater in group varied because L+ is the sole predictor of the food's location. In the fixed group, in contrast, the additional cues provided by the directionally polarizing curtains would share or even overshadow¹⁴ any spatial learning associated with L+.

Learning to approach F+ was characterized by more direct paths in group fixed than in group varied (Fig. 1*b, c*). The crucial data of interest were obtained during a series of post-training transfer tests. With the landmark and directional cues present (but feeders removed), the paths taken by all rats in each group (Fig. 2*a, b*) showed a striking dissociation: search was concentrated at the former location of F+ in group fixed but around L+ in group varied. This suggests that learning where the food was hidden was restricted to group fixed. But this test examines

control of search location by both sets of cues and the result would be trivial if the context cues are of primary importance. A second transfer test was therefore done with the white curtains present but L+ and L- absent. No localized search behaviour at F+ was observed in either group (Fig. 2*c, d*). We calculated the degree of control by cues other than the context cues by subtracting the pixel matrices of Fig. 2*d* from those of Fig. 2*b*, concentrating on time spent searching at F+ or F- and around L+ or L-. True spatial learning was restricted to group fixed (Fig. 3*a*). Group varied discriminated effectively between L+ and L-, but they directed their search at L+ itself and not to the F+ place nearby (Fig. 3*b*). Definitive evidence that group fixed was not guided to F+ by a local olfactory cue was provided by a final 'two-identical landmark' test¹⁵ (Fig. 4).

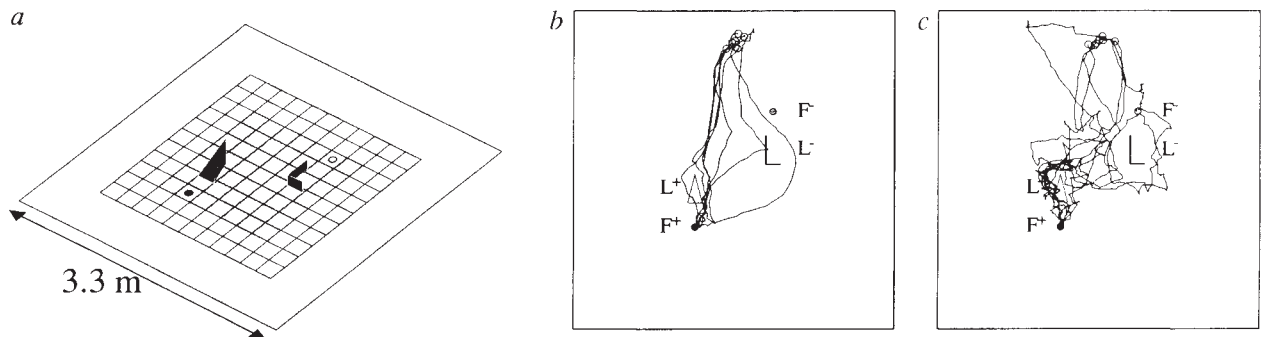
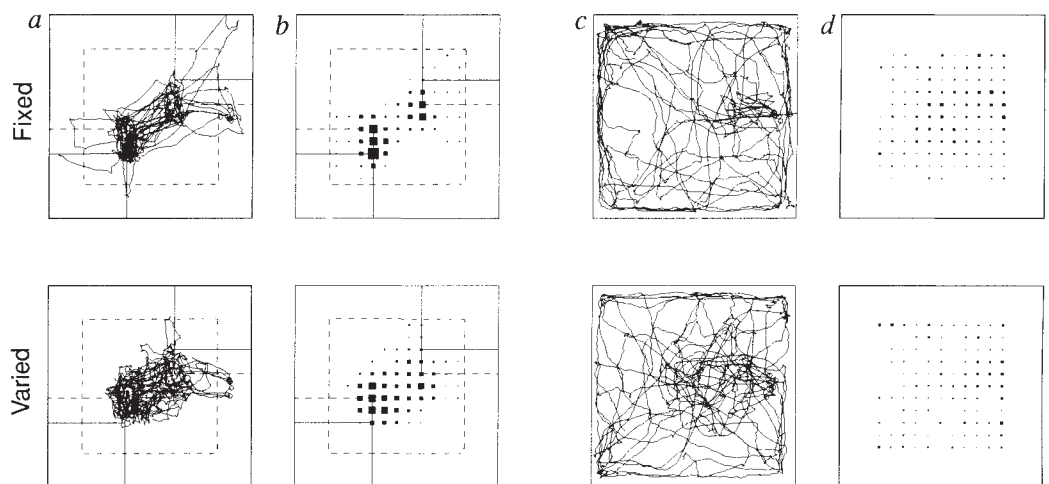


FIG. 1 *a*, Perspective view of training arena. The animals were trained to find food reward in a large square arena (10.89 m²) containing sawdust (5 cm depth) and surrounded by a set of 1 white and 3 black curtains (not shown) and set within a perfectly square room with walls painted black. The set of curtains were rotated 90° daily so that directional polarization was under experimental control. Feeders (5 cm vertical tubes on a movable base) were placed at various locations within the inner 49 pixel grid (20 cm × 20 cm each, shaded grey) and analyses of time spent per pixel were conducted throughout the outer 121 pixel grid (white). The F+ feeder contained accessible food (in this case a breakfast cereal) on top of a 4.2-cm-high dowling insert; the same quantity of reward was inaccessible at the bottom of the F- feeder. Each feeder was placed 40 cm to the 'south' or 'north' of either of two prominent landmarks: (1) an irregular tetrahedron (25 cm high by 30 cm long), covered in sandpaper and marked daily with 1 ml strawberry odour; and (2) an L-shaped wall (20 cm high by 30 cm long), covered in carpet material and marked daily with coffee essence. There were two identical copies of each object that were interchanged regularly between trials. Each object served, for some rats, as L+ (always to the 'north' of

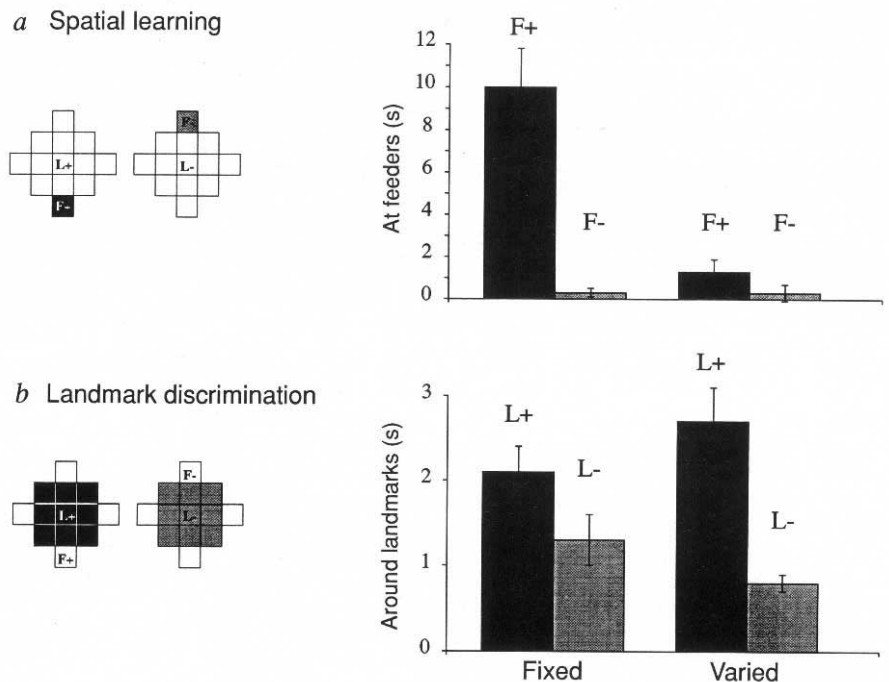
F+), or, for others, as L- (always to the 'south' of F-). The L+F+ pair and the L-F- pair were in fixed locations relative to the white curtains for group fixed (as shown), but in any of 49 locations around the inner pixel grid for group varied. For any placement of the feeders, the closest wall was at a distance of at least 1.05 m and the closest corner 1.48 m. *b* And *c*, representative paths on training trials. There were four rewarded trials and one non-rewarded trial per day. A rewarded training trial consisted of transporting a rat in a slowly rotating opaque box from an adjoining room and putting it into the arena at and facing the centre of the N, S, E or W side-wall (shown as an unfilled circle) from which it could move around freely until it found the F+ feeder; a non-rewarded trial lasted 60 s and was conducted without either landmarks or feeders. The animals' paths were tracked automatically using a hidden overhead videocamera connected to an image-analysis system and an Archimedes computer. The paths shown are those of the 7 rats in group fixed (*b*) and the 7 in group varied (*c*) on trial 85 when the landmark positions in the arena happened to be identical for the two groups. Note that the paths are less direct in group varied.

FIG. 2 *a*, Paths during the transfer test with landmarks. All rats (paths superimposed) searched for the now absent feeder in the central region. Group fixed (above) concentrated its search at the correct feeder location (intersection of uninterrupted lines in the lower left region), while group varied searched randomly around the L+ landmark (intersection of uninterrupted and dotted lines). *b*, Pixel time plots for transfer test with landmarks. The plots (cumulative time per pixel) are normalized such that all pixels are scaled in relation to the pixel with the maximum cumulative time (the F+ pixel of group fixed); the square dotted line represents the extent of the 121 pixel grid. *c* And *d* paths and pixel plots during the transfer test without landmarks. L+ and L- were removed, but the directional curtain cues remained. Both groups patrolled around the



side-walls, spending less time in the central region than when the landmarks were present (time in central region with landmarks present versus with landmarks absent: ANOVA $F=155.3$, d.f. 1/12, $P<0.0001$).

FIG. 3 Spatial learning and landmark discrimination. The pixel matrices of Fig. 2d were first subtracted from those of Fig. 2b; quantitative analysis then focused on selected pixels as indicated in the schematic diagrams. *a*, Spatial learning. An analysis of time spent in the F+ and F- pixels (shaded) revealed a statistical interaction with groups ($F=19.9$, d.f. 1/12, $P<0.001$). Subsequent analysis revealed that only group fixed spent relatively more time at F+ than at F- ($F=49.5$, $P<0.0001$). *b*, Landmark discrimination. An analysis of variance of time spent around L+ and L- (shaded pixels) revealed no difference between groups ($F<1$), greater time around L+ than L- ($F=19.4$, d.f. 1/12, $P<0.001$) and no interaction ($F=3.0$, d.f. 1/12, $P>0.10$). Subsequent analyses showed that the landmark discrimination between L+ and L- was highly significant in group varied ($F=18.8$, d.f. 1/12, $P<0.001$) but did not quite reach significance in group fixed ($F=3.6$, d.f. 1/12, $0.10>P>0.05$). Mean times in seconds $\pm$ 1 s.e.m.



Our finding that geometric stability is a prerequisite for a landmark being used to represent the position of hidden food and that the mere scheduling of a predictable spatial relationship between a moving landmark and hidden food is insufficient is an intriguing dissociation. The lack of spatial learning with moving landmarks suggests that spatial learning processes may have evolved to extract geometric invariance from the environment¹⁶ in a manner analogous to that in which conventional associative learning principles are held to extract the 'causal structure' of the environment¹⁷. As the rats moved around the arena, the relative perceived positions of the landmarks in egocentric space would have been likely to change as much in the fixed as in the varied condition with, for example, L+ being seen as often to the left or right of L- both within and across trials. Constructing a stable representation of the environment from this ever-changing view is the function of spatial learning and it is therefore likely that the inclusion of objects into an animal's representation of allocentric space will depend on their

geometric stability. Our findings suggest that this inclusion is the primary determinant of whether a landmark will be used to predict where something else is located. The rule-of-thumb that 'if it moves, don't use it as a landmark' does not prevent a moving object from acquiring associative strength (as a century of maze studies with randomly placed discriminative cues has long established). Instead, associative strength becomes attached to the object and can, in the present case, only trigger approach or avoidance of the landmarks themselves (the phenomenon of autoshaping¹⁸). The dissociation further implies that rats can detect relative spatial proximity even under circumstances in which they do not represent it in memory.

Parametric manipulation of landmark stability offers an intriguing way of enabling or disabling the process of spatial representation and, thus, a principled basis for exploring further whether the hippocampal formation¹⁹ or hippocampal long-term potentiation²⁰ plays a role in the flexible encoding of spatial relationships. Whether the principle of stability can be extended beyond the spatial domain in relation to the structure of memory within mental models^{21,22} remains to be established. □

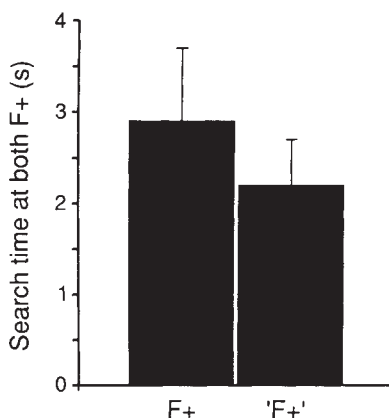


FIG. 4 Two-identical landmark test. All rats were given a further transfer test with two identical L+ objects on trial 5 of the final day. The white curtain was replaced by an additional black curtain and the second 'L+' placed where L- had been located for the four previous rewarded trials. The feeders were absent. The logic of this control procedure is that the animals should choose randomly between the two F+ locations¹⁵ which they were observed to do (paired *t* test, $t=1.06$, d.f. 6, $P>0.3$). Mean time in s $\pm$ 1 s.e.m.

Received 19 November; accepted 14 December 1992.

- O'Keefe, J. *Expl Neurol.* **51**, 78–109 (1976).
- Taube, J. S., Muller, R. U. & Ranck, J. B. Jr *J. Neurosci.* **10**, 420–435 (1990).
- Krebs, J. R., Sherry, D. F., Healy, S. D., Perry, V. H. & Vaccarino, A. L. *Proc. natn. Acad. Sci. U.S.A.* **86**, 1388–1392 (1989).
- Lipp, H.-P. & Schwegler, H. in *The Hippocampus—New Vistas* (eds Chan-Palay, V. & Kohler, C.) 395–410 (Liss, New York, 1989).
- Smith, M. L. & Milner, B. *Neuropsychologia* **19**, 781–793 (1991).
- Morris, R. G. M., Garrud, P., Rawlins, J. N. P. & O'Keefe, J. *Nature* **297**, 681–683 (1982).
- Bingman, V. P., Ioale, P., Casini, G. & Bagnoli, P. *Behav. Brain Res.* **24**, 147–156 (1987).
- Parkinson, J. K., Murray, E. A. & Mishkin, M. *J. Neurosci.* **8**, 4159–4167 (1988).
- O'Keefe, J. & Nadel, L. *The Hippocampus as a Cognitive Map* (Oxford Univ. Press, Oxford, 1978).
- Worden, R. P. *Hippocampus* **2**, 165–188 (1992).
- O'Keefe, J. & Speakman, A. *Expl Brain Res.* **68**, 1–27 (1987).
- Collet, T. S., Cartwright, B. A. & Smith, B. A. *J. comp. Physiol. A* **158**, 835–851 (1986).
- Mackintosh, N. J. *Conditioning and Associative Learning* (Oxford Univ. Press, Oxford, 1983).
- Diez-Chamizo, V., Sterio, D. & Mackintosh, N. J. *Q. J. exp. Psychol.* **37B**, 235–253 (1985).
- Reid, I. C. & Morris, R. G. M. *Proc. R. Soc. B* **247**, 137–143 (1992).
- Sherry, D. F. & Schacter, D. L. *Psychol. Rev.* **94**, 439–454 (1987).
- Dickinson, A. & Mackintosh, N. J. *A. Rev. Psychol.* **29**, 587–612 (1978).
- Brown, P. L. & Jenkins, H. M. *J. exp. Anal. Behav.* **11**, 1–8 (1968).
- Nadel, L. *Hippocampus* **1**, 221–229 (1991).
- Morris, R. G. M. *Cold Spring Harb. Symp. quant. Biol.* **55**, 161–174 (1990).
- Craik, K. *The Nature of Explanation* (Cambridge Univ. Press, Cambridge, 1943).
- Johnson-Laird, P. N. *Mental Models* (Cambridge Univ. Press, Cambridge, 1983).

ACKNOWLEDGEMENTS. We thank R. Spooner for software development and M. Good for discussion. This work was supported by grants from the Medical Research Council, the McDonnell-Pew Institute for Cognitive Neuroscience (University of Oxford) and the Wellcome Trust.

Imaging terminals of *Aplysia* sensory neurons demonstrates role of enhanced Ca^{2+} influx in presynaptic facilitation

L. S. Eliot*, E. R. Kandel, S. A. Siegelbaum† & H. Blumenfeld

Center for Neurobiology and Behavior, College of Physicians and Surgeons of Columbia University, and Howard Hughes Medical Institute, 722 West 168th Street, New York, New York 10032, USA

MODULATION of transmitter release underlies several forms of learning-related synaptic plasticity, including presynaptic facilitation and long-term potentiation¹⁻⁴. Although the presynaptic terminals of most neurons are not accessible for direct study, it has often been possible to correlate changes in calcium influx in the cell body, owing to modulation of K^+ or Ca^{2+} channels, with changes in release⁵⁻⁷. Some forms of presynaptic plasticity, however, do not involve changes in Ca^{2+} influx⁸⁻¹². Moreover, the presence of multiple types of K^+ and Ca^{2+} channels with different subcellular distributions makes the direct measurement of Ca^{2+} influx into presynaptic terminals essential. Using synapses reconstituted in culture between *Aplysia* sensory and motor neurons¹³, we have imaged Ca^{2+} influx in presynaptic terminal regions in response to action potentials, and demonstrate that presynaptic facilitation produced by 5-hydroxytryptamine involves enhanced Ca^{2+} influx through dihydropyridine (DHP)-insensitive Ca^{2+} channels¹⁴ present near release sites. This increased influx is attributable to spike broadening and is significantly correlated with the magnitude of presynaptic facilitation. By contrast, DHP-sensitive channels appear to aid the recovery from depression due to high-frequency stimulation.

Serotonin (5-hydroxytryptamine, 5-HT) increases Ca^{2+} influx into sensory neurons during presynaptic facilitation by two distinct mechanisms. First, 5-HT directly enhances the magnitude of a DHP-sensitive Ca^{2+} current similar to the L-type Ca^{2+} current of vertebrates, without affecting dihydropyridine-insensitive current¹⁵. Thus, as shown in Fig. 1A, 5-HT dramatically increases Ca^{2+} current under voltage clamp (to $220 \pm 36\%$ of control, mean $\pm$ s.e.m., $n=9$). This increase is almost completely blocked (by $89 \pm 3\%$) by the dihydropyridine antagonist nitrendipine ($10 \mu\text{M}$) (Fig. 1A). Despite the increase produced by 5-HT, Ca^{2+} current through L-type channels does not appear to contribute to either normal release or to presynaptic facilitation as the dihydropyridine nifedipine has no effect on either process¹⁴. Presynaptic facilitation has therefore been proposed to result from a second mechanism, whereby 5-HT enhances Ca^{2+} influx indirectly by increasing the duration of the action potential because of closure of K^+ channels¹⁶⁻¹⁹. However, whereas increased spike duration should enhance influx through both populations of Ca^{2+} channels, the data of Fig. 1 and those of Edmonds *et al.*¹⁴ indicate that only influx through the DHP-insensitive channels should contribute to presynaptic facilitation.

To test this idea directly we have used Ca^{2+} imaging and addressed three questions: does 5-HT enhance Ca^{2+} influx during action potentials in presynaptic terminal regions? What are the relative roles of DHP-sensitive and DHP-insensitive Ca^{2+} channels in presynaptic terminals for both normal influx and influx enhanced by 5-HT? Is increased influx through DHP-insensitive Ca^{2+} channels in the terminals adequate to account for presynaptic facilitation?

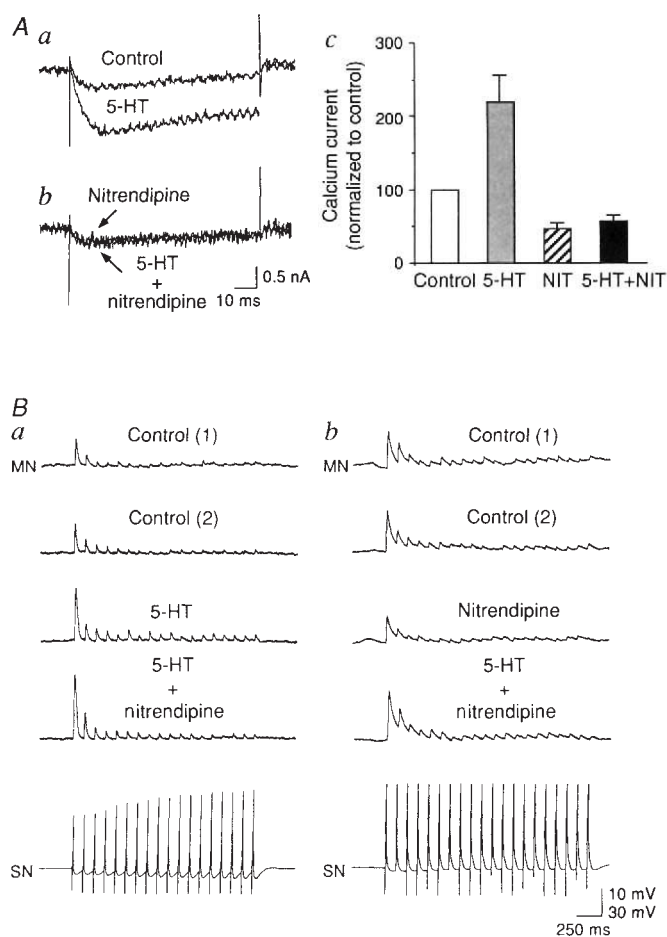


FIG. 1 A, Direct enhancement of nitrendipine-sensitive Ca^{2+} current by 5-HT; Ca^{2+} currents elicited by steps from -50 to $+10$ mV in a dissociated sensory neuron: a, 5-HT ($15 \mu\text{M}$) enhances the Ca^{2+} current; b, after 5-HT was washed out, nitrendipine ($10 \mu\text{M}$) reduced Ca^{2+} current only slightly, but blocked the enhancement of Ca^{2+} current by subsequent 5-HT application; c, group data from 9 experiments. Steady-state calcium current is expressed as per cent of initial control. 5-HT significantly increased the Ca^{2+} current in the absence of nitrendipine ($t=6.04$, $P<0.001$) but not in its presence ($t=1.99$, n.s.). Nitrendipine block in the text is calculated as: (5-HT difference current in control minus 5-HT difference current in nitrendipine)/(5-HT difference current in control). Note that block by DHPs in *Aplysia* is voltage-independent^{14,29}, so concentrations of DHPs that block Ca^{2+} influx through L-type channels under voltage clamp should also block Ca^{2+} influx during action potentials. B, Effect of 5-HT and nitrendipine on e.p.s.p.s elicited by action potential trains: a, intracellular stimulation of a sensory neuron (SN) at 10 Hz for 2 s produces an e.p.s.p. in the motor neuron (MN) that depresses profoundly during the train (control 1); 3 min later, the initial e.p.s.p. has fully recovered (control 2). 5-HT ($15 \mu\text{M}$) facilitates e.p.s.p.s throughout the train. Nitrendipine ($10 \mu\text{M}$) does not reduce facilitation. b, Nitrendipine inhibits recovery from the high-frequency depression following two control trains. 5-HT still facilitates e.p.s.p.s in the presence of nitrendipine. Note that $10 \mu\text{M}$ nitrendipine does not alter synaptic transmission with single action potentials at low frequency (once every 15 min): e.p.s.p. amplitude in nitrendipine was $84 \pm 8\%$ of control compared with $79 \pm 5\%$ following vehicle alone (0.02 – 0.1% dimethylsulphoxide) ($t(51)=0.57$, n.s.; data not shown).

METHODS. Calcium currents were measured in the whole-cell patch clamp configuration as described¹⁴, except we used discontinuous single electrode voltage clamp on an Axoclamp 2A amplifier (switching frequency 20–40 kHz). The pipette solution contained 450 mM CsCl, 2 mM MgCl_2 , 50 mM Cs-HEPES, 10 mM Cs-BAPTA, 5 mM Na_2ATP , 0.1 mM GTP, and 10 mM reduced glutathione, pH 7.2. The bath solution contained 460 mM TEA-Cl, 11 mM CaCl_2 , 55 mM MgCl_2 , 10 mM CsCl, 0.1 mM 3,4-diaminopyridine, and 10 mM Na-HEPES, pH 7.6. Linear leakage and capacitive currents were subtracted. Intracellular recording from sensory and motor neurons is described in Fig. 2 legend. We used nitrendipine (gift from Miles Pharmaceuticals) rather than nifedipine to block L-type Ca^{2+} channels because of its superior stability under the ultraviolet illumination used in imaging experiments.

* Present address: Division of Neuroscience, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA.

† To whom correspondence should be addressed.

To elicit sufficiently large Ca^{2+} signals for imaging, we stimulated presynaptic sensory neurons with trains of action potentials (10 Hz, 2 s), as shown in Fig. 1b. During a train, the excitatory postsynaptic potential (e.p.s.p.) in the motor neuron depresses rapidly. This rapid, 'high-frequency' depression differs from low-frequency depression, which contributes to behavioural habituation, in that it reverses rapidly and is largely recovered after 3 min of rest²⁰.

In response to this train of action potentials, there is a transient elevation of intracellular free Ca^{2+} levels in presynaptic terminal regions of the sensory neurons, as is evident in the Fura-2 ratio image^{21,22} (Fig. 2Aa). 5-HT significantly enhances both the Ca^{2+} transient and the amplitude of the associated postsynaptic potentials throughout the train (Figs 1Ba, 2Ab and Cb). On average, 5-HT increases the Ca^{2+} transient to $143 \pm 11\%$ of its control level ($n = 5$; $P < 0.005$ by paired *t*-test) and increases the size of the initial postsynaptic potential to $186 \pm 28\%$ of control ($P < 0.05$; Fig. 4a). This modulation of the Ca^{2+} transient in presynaptic terminal regions extends to the synaptic region of the previous finding that 5-HT modulates Ca^{2+} influx during action potentials in sensory neuron cell bodies and growth cones²³.

To what extent does this enhanced Ca^{2+} influx occur through the DHP-insensitive channels responsible for release? To isolate this component of Ca^{2+} influx, we blocked Ca^{2+} influx via L-type

channels with 10 μM nitrendipine. Applying nitrendipine in the maintained presence of 5-HT causes some decrease in the Ca^{2+} transient, although it remains elevated relative to its initial pre-5-HT control level (Fig. 2Ac). Moreover, 10 μM nitrendipine does not reduce the size of the e.p.s.p. in the presence of 5-HT (Figs 1B and 2Cc), confirming that the L-type Ca^{2+} current does not contribute to facilitation¹⁴. In the presence of nitrendipine and 5-HT, the Ca^{2+} transient was elevated to $123 \pm 9\%$ ($n = 5$) of the initial control while the e.p.s.p. was facilitated to $196 \pm 60\%$ of control (Fig. 4a).

These results indicate that both types of Ca^{2+} channels are present in synaptic terminal regions and contribute to Ca^{2+} influx with 5-HT. However, the 23% enhancement of the Ca^{2+} transient in 5-HT plus nitrendipine underestimates the extent to which 5-HT enhances influx through the DHP-insensitive channels, because the DHP-sensitive channels contribute to Ca^{2+} influx under control conditions. To assess better the extent to which 5-HT increases Ca^{2+} influx through DHP-insensitive channels, we reversed the order of the experiment, first blocking L-type channels with nitrendipine and then applying 5-HT (Fig. 3).

Nitrendipine alone produced a small decrease in the Ca^{2+} transient to $84 \pm 4\%$ ($n = 5$; $P < 0.01$) of control (Figs 3A, 4a), confirming that L-type channels contribute, albeit modestly, to Ca^{2+} influx during the action potential in the absence of 5-HT.

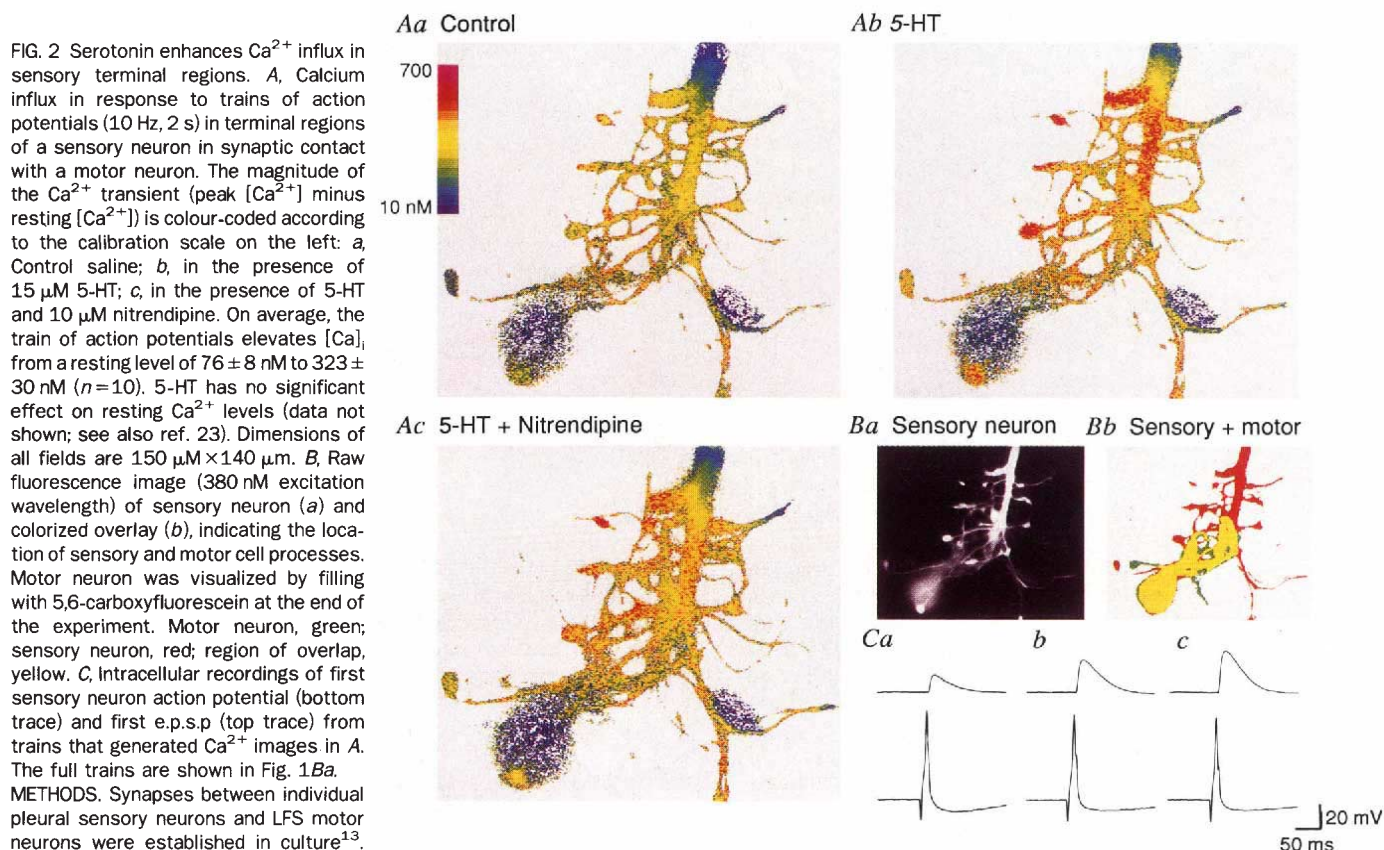


FIG. 2 Serotonin enhances Ca^{2+} influx in sensory terminal regions. A, Calcium influx in response to trains of action potentials (10 Hz, 2 s) in terminal regions of a sensory neuron in synaptic contact with a motor neuron. The magnitude of the Ca^{2+} transient (peak $[\text{Ca}^{2+}]$ minus resting $[\text{Ca}^{2+}]$) is colour-coded according to the calibration scale on the left: a, Control saline; b, in the presence of 15 μM 5-HT; c, in the presence of 5-HT and 10 μM nitrendipine. On average, the train of action potentials elevates $[\text{Ca}]_i$ from a resting level of 76 ± 8 nM to 323 ± 30 nM ($n = 10$). 5-HT has no significant effect on resting Ca^{2+} levels (data not shown; see also ref. 23). Dimensions of all fields are $150 \mu\text{m} \times 140 \mu\text{m}$. B, Raw fluorescence image (380 nm excitation wavelength) of sensory neuron (a) and colorized overlay (b), indicating the location of sensory and motor cell processes. Motor neuron was visualized by filling with 5,6-carboxyfluorescein at the end of the experiment. Motor neuron, green; sensory neuron, red; region of overlap, yellow. C, Intracellular recordings of first sensory neuron action potential (bottom trace) and first e.p.s.p. (top trace) from trains that generated Ca^{2+} images in A. The full trains are shown in Fig. 1Ba. METHODS. Synapses between individual pleural sensory neurons and LFS motor neurons were established in culture¹³. We estimate that 2–5% of the presynaptic membrane area is occupied by active zones³⁰. Experiments were performed 2–3 days after plating. Most of region of sensory-motor neuron contact fits into a single field of view using $40\times$ objective. Intracellular Ca^{2+} was measured as described²³, except that an intensified CCD camera (Hamatsu) was the detector. Ratio images (340/380 nm excitation wavelengths) were generated on line every 2 s from the average of 8 background-subtracted frames at each wavelength and then stored to disk using a Videoprobe real-time imaging device (ETM Systems). Sensory neurons were filled by iontophoresis with 10 mM Fura-2 pentapotassium salt (Molecular Probes) in 0.5 M potassium acetate, 10 mM HEPES, pH 7.2. The same microelectrode (30–40 Mohm) was used to elicit

and record action potentials with a 2-s train of 20 brief (2–5 ms) depolarizing pulses. Trains were elicited at 3-min intervals, and 2 trains were delivered in each drug condition. The motor neuron was impaled with a microelectrode containing 2 M potassium acetate, 0.5 M KCl (15–25 M Ω resistance), and hyperpolarized to -80 mV to measure e.p.s.p.s. Experiments were done at 18°C in supplemented L15 medium¹³. Drugs were applied using a multichannel microperfusion system. Because nitrendipine was prepared as a 10 mM stock in ethanol, 0.1% ethanol was included in all solutions; it had no obvious effect on the electrophysiological properties of the cells.

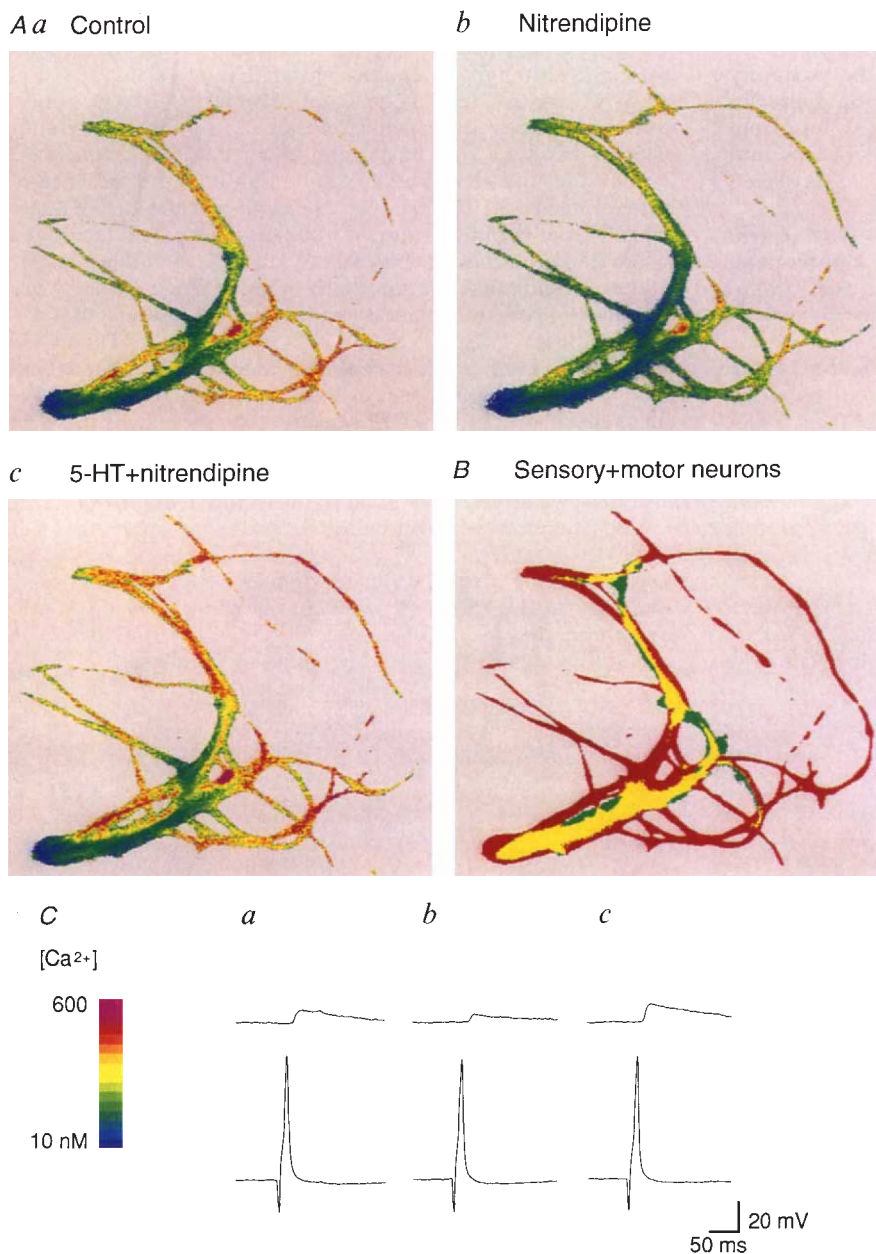


FIG. 3 Serotonin increases Ca^{2+} influx in terminal regions in the presence of nitrendipine. **A**, Presynaptic Ca^{2+} transients (peak minus resting concentration) to a train of 20 action potentials under control conditions (**a**) in the presence of $10 \mu\text{M}$ nitrendipine (**b**), and in the presence of nitrendipine and $15 \mu\text{M}$ 5-HT (**c**). **B**, Location of sensory and motor neuron processes. Same colour code as for Fig. 2Bb. **C**, The first sensory neuron action potential and motor neuron e.p.s.p. of the trains that elicited the images in **A**. Dimensions of all fields are $160 \mu\text{m} \times 140 \mu\text{m}$.

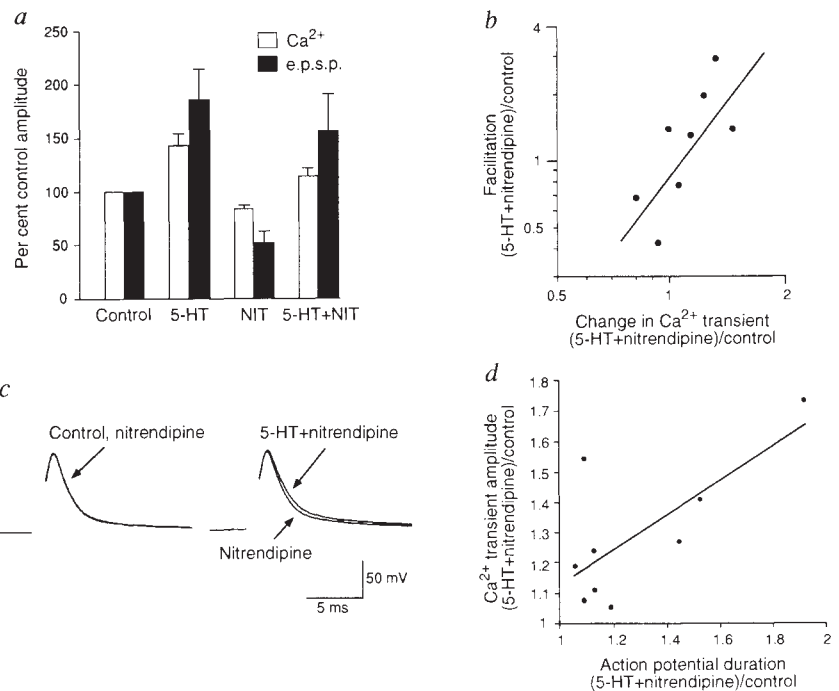
Surprisingly, nitrendipine also reduced e.p.s.p. amplitude (to $52 \pm 11\%$ of control). This decrease is significant compared with the normal amount of depression in e.p.s.p. amplitude that occurs from one control train to the next ($81 \pm 7\%$; $t(9) = 2.42$, $P < 0.05$; Fig. 1Bb). This effect of nitrendipine on e.p.s.ps elicited by repeated trains contrasts with the finding that dihydropyridines do not alter e.p.s.p. amplitude when elicited by single action potentials at low frequency¹⁴ (see legend to Fig. 1). These results suggest that although L-type channels do not directly mediate release to a single action potential, they may modulate^{9,14} release by promoting recovery from high-frequency depression.

Despite this effect of nitrendipine on control e.p.s.ps, 5-HT still produced normal presynaptic facilitation and significantly enhanced the Ca^{2+} transient when applied in the presence of nitrendipine (Figs 1Bb, 3Ac and Cc). On average, 5-HT increased the Ca^{2+} transient to $131\% \pm 9\%$ ($n = 5$; $P < 0.01$) of its value in nitrendipine alone (Fig. 4). Under these same conditions, 5-HT facilitated the synaptic potential to $216\% \pm 63\%$ of its amplitude in the presence of nitrendipine alone, consistent with the L-type channels not being required for facilitation (Fig. 4).

This increase in Ca^{2+} influx through DHP-insensitive channels seems likely to contribute significantly to presynaptic facilitation because the two processes were significantly correlated in the presence of nitrendipine ($R = 0.74$, $P < 0.05$). As shown in Fig. 4b, facilitation varied with the enhanced Ca^{2+} signal logarithmically by a power of 2.4, consistent with the cooperative binding of two or more Ca^{2+} ions to trigger exocytosis of a single vesicle^{24,25}. Although we cannot directly measure action potential duration in the presynaptic terminals, parallel experiments in the cell body show a direct relationship between the increase in Ca^{2+} influx and the increase in spike duration with 5-HT (Fig. 4c, d). As 5-HT has no direct effect on the DHP-insensitive Ca^{2+} current in the cell body, these results suggest that the DHP-insensitive increase in Ca^{2+} influx we measure in terminal regions is a consequence of the broadening of the presynaptic action potential.

These results provide direct evidence that increased Ca^{2+} influx contributes to the presynaptic facilitation produced by 5-HT in *Aplysia* sensory neurons. They also give further insight into the functional roles of different Ca^{2+} channels. Ca^{2+} influx through DHP-insensitive channels appears to trigger release directly and increasing this influx by broadening the action

FIG. 4 Data summary. *a*, Average results showing effects of nitrendipine ($n=5$), 5-HT ($n=5$), and nitrendipine + 5-HT ($n=10$), on the Ca^{2+} transient and e.p.s.p. Data for 5-HT + nitrendipine were pooled from experiments using both protocols depicted in Figs 2 and 3. To measure amplitude of Ca^{2+} transients, resting Ca^{2+} values were subtracted from peak Ca^{2+} values for all pixels in the field whose fluorescence intensity was greater than background. Calcium transient values were then averaged for all non-zero pixels. *b*, Relationship between changes in Ca^{2+} influx and synaptic facilitation by 5-HT in individual experiments. Data are normalized to control so that experiments using two different protocols (see Figs 2 and 3) could be pooled. The number of trials between control and 5-HT + nitrendipine was the same in both protocols. To relate the change in presynaptic Ca^{2+} to transmitter release, we have compared the magnitude of the Ca^{2+} transient to the amplitude of the first e.p.s.p. of the train. The Ca^{2+} transient should be proportional to Ca^{2+} influx during the first action potential, because the amplitude of the Ca^{2+} transient varies linearly with the number of action potentials in the sensory neurons²². *c*, Action potential broadening by 5-HT in the presence of nitrendipine. The first action potential of trains delivered in control, 10 μM nitrendipine, and 15 μM 5-HT + 10 μM nitrendipine are superimposed. 5-HT in the presence of nitrendipine increased the somatic Ca^{2+} transient by an average of $31 \pm 9\%$ ($n=5$, $P < 0.05$) and the somatic action potential duration by $29 \pm 9.7\%$ ($n=9$, $P < 0.02$). *d*, Relationship between spike broadening and somatic Ca^{2+} influx in individual experiments. Action potential duration refers to the time between peak and 50%



repolarization for the first action potential of the train. Similar results were obtained using the last action potential of train.

potential enhances release. By contrast, influx through DHP-sensitive channels plays a modulatory role in maintaining the level of release during extensive activity.

The enhancement of Ca^{2+} influx through DHP-insensitive channels is likely to be an important mechanism by which 5-HT increases release at non-depressed and moderately depressed synapses. But at synapses that are more extensively depressed by low-frequency stimulation, another mechanism, independent of Ca^{2+} influx, is thought to contribute to presynaptic facilitation^{9,10,26-28}. Together, these data suggest that both modulation of Ca^{2+} influx during the action potential and changes independent of Ca^{2+} influx operate at a single synapse and their relative contribution may vary as a function of a neuron's previous activity. □

Phosphorylation and modulation of recombinant GluR6 glutamate receptors by cAMP-dependent protein kinase

Lynn A. Raymond*, Craig D. Blackstone†
& Richard L. Huganir†‡

† Department of Neuroscience, * Department of Neurology,
Howard Hughes Medical Institute, The Johns Hopkins University
School of Medicine, Baltimore, Maryland 21205, USA

Received 19 October; accepted 15 December 1992.

- Hawkins, R. D., Kandel, E. R. & Siegelbaum, S. A. *Rev. Neurosci.* (in press).
- Castellucci, V. F. & Kandel, E. R. *Science* **194**, 1176-1178 (1976).
- Bekkers, J. M. & Stevens, C. F. *Nature* **346**, 724-729 (1990).
- Malinow, R. & Tsien, R. W. *Nature* **346**, 177-180 (1990).
- Dunlap, K. & Fischbach, G. D. *Nature* **276**, 837-839 (1978).
- Klein, M. & Kandel, E. R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6912-6916 (1980).
- Lipscombe, D., Kongsamut, S. & Tsien, R. W. *Nature* **340**, 639-642 (1989).
- Man-Son-Hing, H., Zoran, M. J., Lukowiak, K. & Haydon, P. G. *Nature* **341**, 237-239 (1989).
- Dale, N. & Kandel, E. R. *J. Physiol., Lond.* **421**, 203-222 (1990).
- Hochner, B., Klein, M., Schacher, S. & Kandel, E. R. *Proc. natn. Acad. Sci. U.S.A.* **83**, 8794-8798 (1986).
- Delaney, K., Tank, D. W. & Zucker, R. S. *J. Neurosci.* **11**, 2631-2643 (1991).
- Malgaroli, A. & Tsien, R. W. *Nature* **357**, 134-139 (1992).
- Rayport, S. G. & Schacher, S. *J. Neurosci.* **6**, 759-763 (1986).
- Edmonds, B., Klein, M., Dale, N. & Kandel, E. R. *Science* **250**, 1142-1147 (1990).
- Braha, O., Edmonds, B., Klein, M., Sacktor, T. & Kandel, E. R. *J. Neurosci.* (in press).
- Hochner, B., Klein, M., Schacher, S. & Kandel, E. R. *Proc. natn. Acad. Sci. U.S.A.* **83**, 8410-8414 (1986).
- Siegelbaum, S. A., Camardo, J. S. & Kandel, E. R. *Nature* **299**, 413-417 (1982).
- Baxter, D. A. & Byrne, J. H. *J. Neurophysiol.* **62**, 665-679 (1989).
- Mercer, A. R., Emptage, N. J. & Carew, T. J. *Science* **254**, 1811-1813 (1991).
- Byrne, J. H. *J. Neurophysiol.* **48**, 431-438 (1982).
- Grynkiewicz, G., Poenie, M. & Tsien, R. Y. *J. Biol. Chem.* **260**, 3440-3450 (1985).
- Blumenfeld, H., Zablow, L. & Sabatini, B. *Biophys. J.* **63**, 1146-1164 (1992).
- Blumenfeld, H., Spira, M. E., Kandel, E. R. & Siegelbaum, S. A. *Neuron* **5**, 487-499 (1990).
- Dodge, F. A. & Rahamimoff, R. *J. Physiol., Lond.* **193**, 419-432 (1967).
- Smith, S. J., Augustine, G. J. & Charlton, M. P. *Proc. natn. Acad. Sci. U.S.A.* **82**, 622-625 (1985).
- Braha, O. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 2040-2044 (1990).
- Gingrich, K. J. & Byrne, J. H. *J. Neurophysiol.* **53**, 652-669 (1985).
- Ghirardi, M. *et al. Neuron* **9**, 479-489 (1992).
- Nerbonne, J. M. & Gurney, A. M. *J. Neurosci.* **7**, 882-893 (1987).
- Glanzman, D. L., Kandel, E. R. & Schacher, S. *Neuron* **3**, 441-450 (1989).

ACKNOWLEDGEMENTS. We thank B. Edmonds for help with initial experiments, S. Mack and C. Lamb for preparing the figures, and H. Ayers and A. Krawetz for typing the manuscript.

GLUTAMATE-GATED ion channels mediate most excitatory synaptic transmission in the central nervous system and play crucial roles in synaptic plasticity, neuronal development and some neuropathological conditions¹⁻³. These ionotropic glutamate receptors have been classified according to their preferred agonists as NMDA (*N*-methyl-D-aspartate), AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionate) and KA (kainate) receptors⁴. On the basis of sequence similarity and pharmacological properties, the recently cloned glutamate receptor subunits have been assigned as components of NMDA (NMDAR1, 2A-D), AMPA (GluR1-4) and KA (GluR5-7, KA1, KA2) receptors⁵⁻⁷. Protein phosphorylation of glutamate receptors by protein kinase C and cyclic AMP-dependent protein kinase (PKA) has been suggested to regulate their function⁸⁻¹⁸, possibly playing a prominent role in certain forms of synaptic plasticity such as long-term potentiation¹⁹ and long-term depression⁹. Here we report that the GluR6 glutamate receptor, transiently expressed in mammalian cells, is directly phosphorylated by PKA, and that intracellularly applied PKA increases the amplitude of the glutamate response. Site-specific mutagenesis of the serine residue (Ser 684) representing a PKA consensus site completely eliminates PKA-mediated phosphorylation of this site as well as the potentiation of the glutamate

‡ To whom correspondence should be addressed.

response. These results provide evidence that direct phosphorylation of glutamate receptors modulates their function.

To examine whether GluR6 is phosphorylated by PKA, we expressed the rat GluR6 glutamate receptor in human embryonic kidney 293 (HEK-293) cells. Expression of this protein was confirmed by immunoblotting using anti-peptide antibodies directed against GluR6 (anti-R6). The anti-R6 antibodies detected a major band at 112K and a minor band at 99K, as well as a lower molecular mass degradation product (Fig. 1a). After enzymatic deglycosylation of membranes with PNGase F, only the 99K species was detected, suggesting that this band represents an unglycosylated precursor of the mature receptor (Fig. 1a). The anti-R6 antibodies were also able to immunoprecipitate GluR6 specifically from transfected HEK-293 cells prelabelled with [35 S]methionine (Fig. 1b). Using these antibodies, GluR6 was isolated by immunoprecipitation from HEK-293 cells which had been prelabelled with [32 P]orthophosphate, revealing that the GluR6 protein has a low level of basal phosphorylation (Fig. 1c). But on cotransfection with a complementary DNA encoding the catalytic subunit of PKA ($C\alpha$), GluR6 showed a large increase in phosphorylation (Fig. 1c), strongly suggesting that PKA directly phosphorylates GluR6. Phosphoamino acid analysis of the phosphorylated GluR6 protein demonstrated that phosphorylation occurred mostly on serine residues (data not shown; $n=2$). Treatment of GluR6-transfected HEK-293 cells with the adenylyl cyclase activator forskolin (50 μ M) also increased the phosphorylation of the GluR6 protein (data not shown; $n=3$).

The functional effects of the phosphorylation of GluR6 by PKA were examined by analysing the physiological properties of GluR6 expressed in HEK-293 cells. The GluR6 subunit used in our experiments contained an arginine residue in proposed transmembrane domain II (GluR6(R))²⁰. These homooligomeric GluR6 receptors exhibited rapid desensitization to both kainate and glutamate as previously reported²⁰, and the peak current-voltage ($I-V$) relationship was essentially linear (Fig. 3c, Table 1). Application of 200 μ M glutamate to the cells resulted in a rapidly activating inward current that decayed with a time constant of 8–14 ms (Fig. 3a, Table 1). The peak amplitude and desensitization kinetics of this response, using an ATP-containing pipette solution (control), were very stable for periods up to 50 min (Fig. 3a, d, Table 1). The further addition of creatine phosphokinase (CPK) and phosphocreatine to the pipette solution did not affect the stability of the glutamate-evoked response. Omitting ATP from the pipette solution resulted in only minimal rundown; the peak response 30 min after whole-cell seal formation was $98 \pm 11\%$ ($n=4$) of the initial peak current, suggesting that basal phosphorylation of the GluR6 receptor is not required for maintenance of the glutamate-evoked response. The low level of basal phosphorylation seen for GluR6 expressed in HEK-293 cells in the absence of PKA stimulation (Fig. 1c) is consistent with this result.

In contrast, when the purified catalytic subunit of PKA was included in the solution used to backfill the patch pipette, and allowed to perfuse the cell, potentiation of the glutamate-evoked response was observed (Fig. 3b, Table 1) in 12 of 14 cells tested. This potentiation was detectable within 5–10 min and was complete by 35 min (Fig. 3d). Intracellular perfusion of cAMP similarly potentiated the response of the cells to glutamate (data not shown; $n=4$). PKA phosphorylation of the GluR6 protein had no significant effect on the $I-V$ relationship, dose-response, rise time or desensitization kinetics of the glutamate-evoked current (Table 1). The specificity of the PKA effect was demonstrated using PKI_{5–24} amide, a specific peptide inhibitor of PKA. When PKI was included in the patch pipette solution along with PKA, the potentiation was markedly diminished (Table 1). In addition, the effect of PKI appeared to be dose-dependent; at a concentration of 500 μ g ml⁻¹, PKI completely eliminated potentiation (Fig. 3e, h; $n=1$), whereas at 400 μ g ml⁻¹, PKI

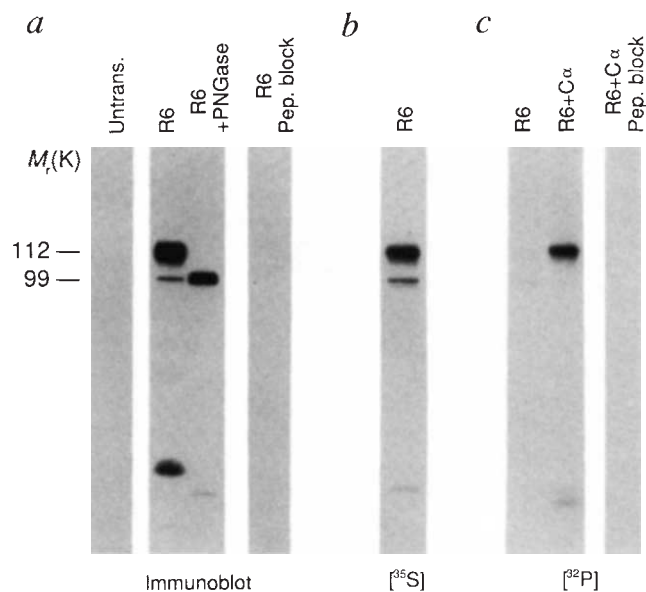


FIG. 1 Phosphorylation of the GluR6 glutamate receptor by PKA. **a**, HEK-293 cells were transiently transfected with a mammalian expression vector containing the rat GluR6 cDNA. Membranes from these cells (20 μ g protein per lane) were immunoblotted with anti-R6 antibodies, as were membranes that had been pretreated with PNGase F to remove *N*-linked glycosylation chains. No immunodetection was observed using membranes from untransfected cells (Untrans.). Immunorecognition of GluR6 was abolished when anti-R6 antibodies were preadsorbed with the synthetic peptide (10 μ g ml⁻¹) to which they were raised. Molecular sizes (in K) are indicated to the left. **b**, GluR6-transfected HEK-293 cells were prelabelled with [35 S]methionine, and the GluR6 protein was isolated from solubilized membranes by immunoprecipitation with anti-R6 antibodies. Immune complexes were resolved by SDS-PAGE and analysed by fluorography. **c**, HEK-293 cells were transfected with GluR6 or cotransfected with GluR6 and $C\alpha$ expression plasmids as indicated. The $C\alpha$ construct encodes the catalytic subunit of PKA. Cells were prelabelled with [32 P]orthophosphate, and GluR6 was isolated from solubilized membranes by immunoprecipitation with anti-R6 antibodies. The immunoprecipitates were subjected to SDS-PAGE, and phosphorylated proteins were revealed by autoradiography. Where indicated, specific immunoprecipitation of GluR6 was blocked by preadsorption of anti-R6 antibodies with the synthetic peptide (10 μ g ml⁻¹).

METHODS. A cDNA clone encoding rat GluR6 (ref. 20) was excised from pBlueScript KS⁺ as an *EcoRI*-*XbaI* fragment and subcloned into a mammalian expression vector under control of the cytomegalovirus promoter. Human embryonic kidney 293 (HEK-293) cells (American Type Culture Collection CRL 1573) were transfected by calcium phosphate coprecipitation with 20 μ g plasmid DNA per 10 cm plate. The $C\alpha$ plasmid, which encodes the catalytic subunit of PKA (3 μ g per plate), was included where indicated. Forty-eight hours after transfection, the cells were labelled with [35 S]methionine (1 mCi ml⁻¹; Tran[35 S]-label, ICN Pharmaceuticals) for 4 h in methionine-free media or with [32 P]orthophosphate (1 mCi ml⁻¹; Dupont Biotechnology Systems) for 4 h in phosphate-free media. After metabolic labelling, the cells were lysed by sonication in 20 mM phosphate buffer (pH 7.0) containing 150 mM NaCl, 50 mM NaF, 1 mM Na₃VO₄ (ortho), 10 mM sodium pyrophosphate, 5 mM EDTA, 5 mM EGTA, and 10 U ml⁻¹ Trasylol (aprotinin), and membranes were collected by centrifugation (114,000g, 10 min, 4 °C). The pellet was resuspended in 1.0% (w/v) SDS, and the extract clarified by centrifugation. Five volumes of ice-cold buffer containing 2.0% Triton X-100 (w/v) were added, and the GluR6 protein was then isolated by immunoprecipitation using affinity-purified rabbit polyclonal antibodies directed against a synthetic peptide (KHTFNDRLPGKETMA; single-letter amino-acid code) corresponding to the C-terminal 15 amino acid residues of the GluR6 protein²⁴ (anti-R6) which had been coupled to Protein A-Sepharose CL-4B (Pharmacia). Immunoprecipitates were resolved by SDS-PAGE (8.0% acrylamide) and either subjected to autoradiography (for 32 P) or incubated with Entensify (Dupont) and analysed by fluorography (for 35 S). Where indicated, membranes were prepared from unlabelled GluR6-transfected HEK-293 cells and treated with or without the glycosidase PNGase F (Boehringer-Mannheim) as previously described²⁵. These membranes were subjected to SDS-PAGE (8.0% acrylamide) and immunoblotted with anti-R6 antibodies (0.5 μ g ml⁻¹), with detection by enhanced chemiluminescence (ECL, Amersham).

TABLE 1 Analysis of the properties of glutamate-evoked currents

Receptor	Treatment	Potential (% initial peak current)	τ_D (ms)	V_{rev} (mV)	$G_{+80/-80}$	EC_{50} (μ M)
GluR6	Control	106 ± 17 (7)	11.7 ± 2.0 (5)	8 ± 2 (6)	1.3 ± 0.2 (6)	273 ± 73 (5)
GluR6	+PKA	155 ± 19 (12)	11.9 ± 1.6 (9)	12 ± 5 (6)*†	1.2 ± 0.2 (6)*†	210 ± 94 (5)*
GluR6	+PKA/PKI	119 ± 7 (4)	11.4 ± 1.9 (3)	—	—	—
GluR6 (S684A)	+PKA	109 ± 14 (7)	11.0 ± 1.9 (6)	13 ± 9 (4)	1.3 ± 0.4 (4)	—

Current responses to rapid application of 200 μ M glutamate were recorded from HEK-293 cells under whole-cell voltage clamp. Values represent the mean $\pm$ s.d., the number of cells recorded for each measurement is indicated in parentheses. Recordings were made using a control pipette solution (Fig. 3a) or with pipette solution containing the purified catalytic subunit of PKA (20–30 μ g ml⁻¹). Dashes indicate properties not analysed. Potentiation is the peak amplitude of the glutamate current recorded 20–35 min after whole-cell seal formation expressed as a percentage of the initial current. τ_D is the decay time constant, which in every case was fitted well by a single exponential; V_{rev} is the reversal potential; $G_{+80/-80}$ is the ratio of conductances at holding potentials of +80 mV and -80 mV; EC_{50} is the glutamate concentration at which the current is half maximal. The rise times (10–90%) of the glutamate-evoked currents were (mean $\pm$ s.d., in ms): 7.7 ± 1.1 (GluR6, control, $n=5$); 7.4 ± 1.3 (GluR6, +PKA, $n=9$); and 7.5 ± 0.8 (GluR6 (S684A), +PKA, $n=6$).

* Some of the cells included were recorded with pipette solution containing 300 μ M cAMP, and one cell was cotransfected with cDNA for Ca.

† The glutamate concentration used was 300 μ M for one cell and 1 mM for another; all other cells were recorded using 200 μ M glutamate.

blocked most, but not all, of the effect of PKA (peak potentiation of $122 \pm 5\%$ of the initial current, $n=3$).

The proposed major intracellular domains of cloned glutamate receptor subunits contain consensus sites for phosphorylation by several protein kinases²¹, but the GluR6 subunit is the only one with a strong consensus site for PKA (Arg-Arg-X-Ser) in this region²⁰ (Fig. 2c). To determine whether the potentiation of the glutamate response by PKA was due to direct phosphorylation of the GluR6 protein, site-specific mutagenesis techniques were used to convert the serine residue within the PKA consensus site (Ser 684) to an alanine residue. The physiological properties of this mutant GluR6 protein were identical to those of the wild-type protein (Fig. 3f, g, Table 1). But in contrast to results obtained with wild-type receptors, intracellular perfusion of purified PKA did not potentiate the glutamate response of the mutant GluR6 receptors (Fig. 3f, h, Table 1). To establish biochemically that Ser 684 was directly phosphorylated by PKA, GluR6 was immunoprecipitated from [³²P]orthophosphate-labelled HEK-293 cells coexpressing wild-type or mutant GluR6 along with the catalytic subunit of PKA (Ca). Site-specific mutagenesis of Ser 684 to an alanine residue resulted in a large

decrease in PKA-mediated phosphorylation of GluR6 (Fig. 2a). Two-dimensional phosphopeptide map analysis²² revealed that wild-type GluR6 is phosphorylated on one major site and several minor sites (Fig. 2b). Mutagenesis of Ser 684 to an alanine residue eliminated phosphorylation of GluR6 on the major phosphopeptide. These results provide strong evidence that PKA directly phosphorylates GluR6 on Ser 684 and that this phosphorylation is required for the potentiation of the glutamate response. Because Ser 684 is contained within the proposed major intracellular domain of the GluR6 protein (Fig. 2c), these data support the proposed transmembrane topology model for GluR6 and other glutamate receptors^{5–7} and confirm that this region is intracellular.

Accumulating evidence suggests that protein phosphorylation is an important mechanism for modulating glutamate receptor function. Activation of protein kinase C (PKC) in neurons modulates the responses of both non-NMDA⁹ and NMDA^{10–12} receptors. Moreover, glutamate receptors gated by kainate in white perch retinal horizontal cells and mammalian hippocampal neurons can be regulated by cAMP-dependent protein phosphorylation^{13–16}. As in our study, the amplitude of

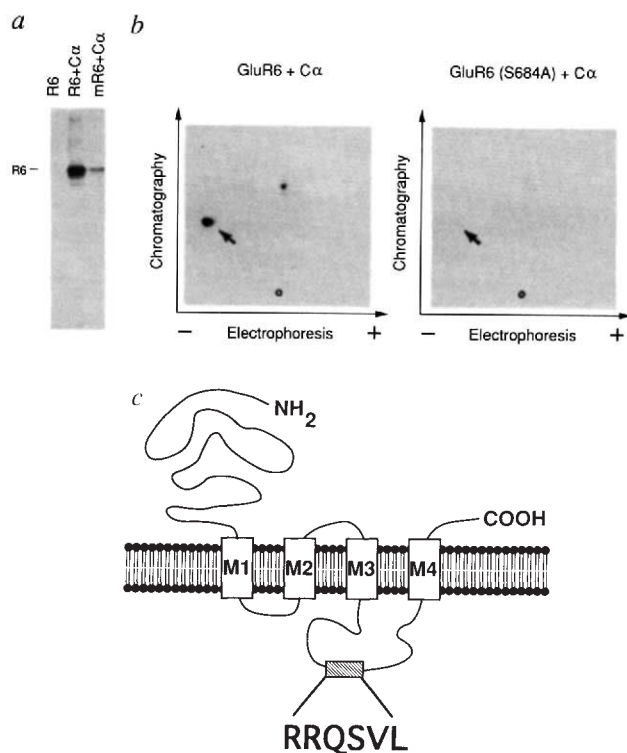


FIG. 2 PKA-mediated phosphorylation of GluR6 on Ser 684. *a*, HEK-293 cells were transfected with expression constructs for wild-type GluR6 (R6) or a mutant GluR6 in which Ser 684 had been replaced with an Ala residue (mR6) by site-specific mutagenesis. Ca was cotransfected where indicated. These cells were prelabelled with [³²P]orthophosphate, and solubilized membranes were prepared. The GluR6 protein was immunoprecipitated with anti-R6 antibodies and resolved by SDS-PAGE. Phosphorylated GluR6 protein was detected by autoradiography. *b*, Phosphopeptide map analysis of PKA phosphorylated GluR6. The wild-type and S684A mutant GluR6 phosphoproteins shown in *a* were excised from the gel, digested with trypsin (300 μ g ml⁻¹), and subjected to two-dimensional phosphopeptide map analysis. Arrows indicate the major phosphopeptide which is not present in maps of the GluR6 (S684A) mutant. Similar results were obtained in four additional experiments. The origin is indicated (○). *c*, Schematic model of proposed membrane topology of GluR6, showing an individual subunit with transmembrane domains M1–M4 viewed through the plane of the membrane. Both the N and C termini are extracellular. The major intracellular domain between M3 and M4 contains Ser 684, the site for phosphorylation by PKA.

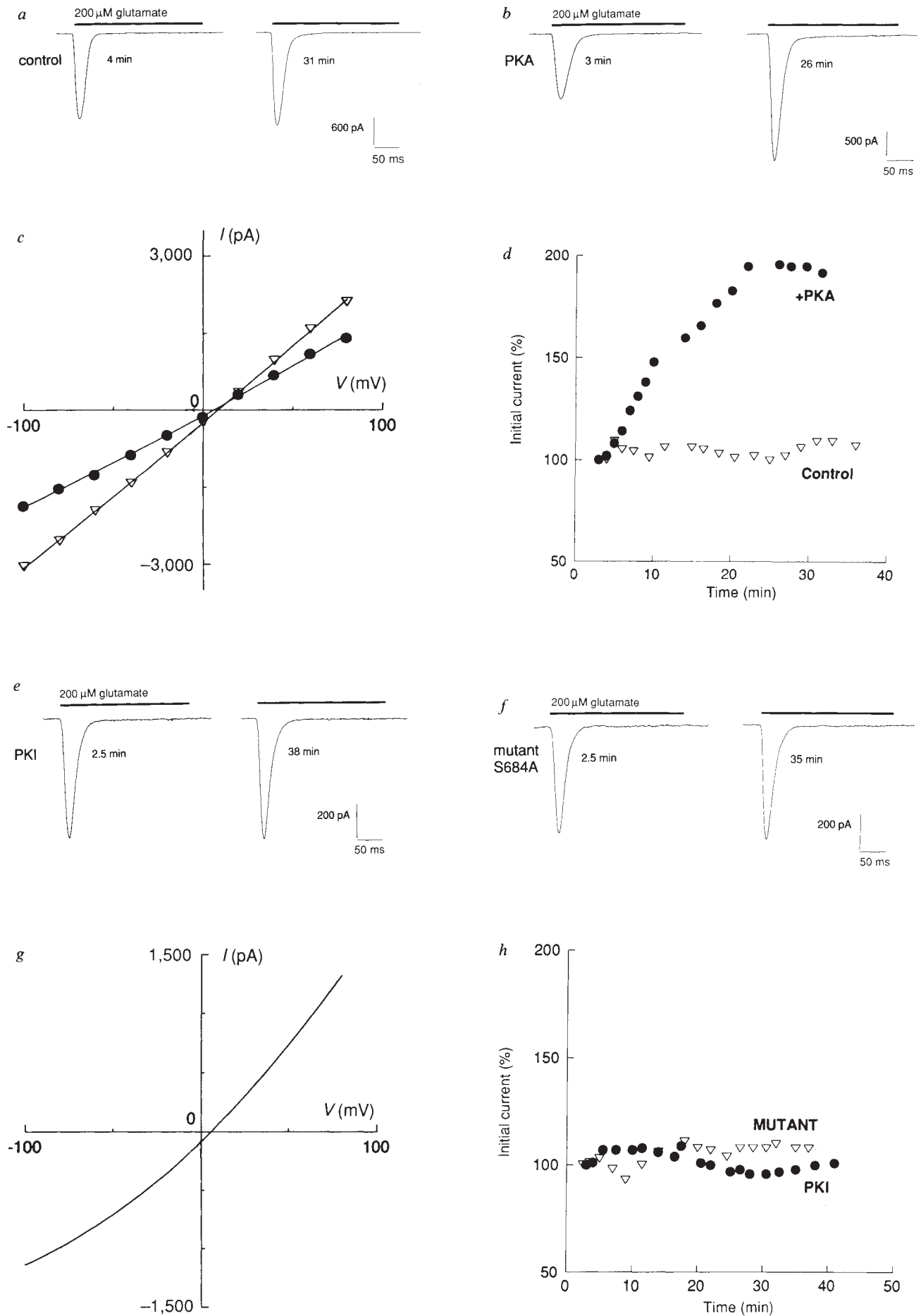


FIG. 3 *a-d*, PKA potentiation of glutamate-induced current in HEK-293 cells transfected with GluR6. *a, b*, Whole-cell membrane currents in response to rapid application of glutamate were recorded under voltage clamp ($V_H = -60$ mV) from two different cells transiently expressing GluR6(R) subunits. The patch pipette was backfilled with the control pipette solution in the recording shown in *a*, whereas $20 \mu\text{g ml}^{-1}$ of freshly diluted, purified catalytic subunit of PKA was added to the pipette backfill solution and allowed to diffuse into the cell shown in *b*. After rupturing the patch to form the whole-cell recording mode and lifting the cell from the recording chamber floor, brief pulses of glutamate were applied at 1 min intervals. The initial current responses, shown to the left, show the glutamate-induced membrane current at 4 min (*a*) or 3 min (*b*) after whole-cell seal formation. The right-hand traces show the responses 31 min (*a*) or 26 min (*b*) after whole-cell seal formation. *c*, Peak current-voltage relationship for GluR6 with and without PKA treatment. The voltage was stepped from -100 to $+80$ mV in 20 mV increments at 1 min intervals. The peak current elicited by a brief pulse of glutamate ($200 \mu\text{M}$) was recorded at each voltage for a control (∇) and PKA-treated ($\bullet$) cell. The reversal potentials and linearity of the $I-V$ curves are nearly identical under the two conditions. The difference in slope is accounted for by the difference in current amplitude of the two cells. *d*, Time course of potentiation of glutamate currents by PKA. The peak glutamate-activated currents in control (∇) and PKA-treated ($\bullet$) cells were recorded at 1 min intervals and expressed as a percentage of the initial glutamate current recorded within $3-4$ min of whole-cell seal formation. The two curves were generated by analysing recordings from the cells shown in *a* and *b*. *e-h*, Block of PKA-mediated potentiation in GluR6-transfected HEK-293 cells by intracellular perfusion of PKI or by mutation of Ser 684. *e*, Whole-cell voltage clamp recordings from a GluR6-transfected HEK-293 cell showing the membrane current response to rapid perfusion of glutamate. The recording pipette was backfilled with the standard ATP-containing solution as well as $30 \mu\text{g ml}^{-1}$ PKA and $500 \mu\text{g ml}^{-1}$ PKI₅₋₂₄ amide. The glutamate-induced current responses recorded at 2.5 min (baseline) and 38 min after formation of the whole-cell recording mode are almost identical, showing that PKI₅₋₂₄ amide blocked the potentiating effect of PKA. *f*, Whole-cell recordings under voltage-clamp from an HEK-293 cell transiently transfected with mutant GluR6 (S684A), illustrating the membrane current response to rapid glutamate application. The backfill pipette solution was

the standard ATP-containing solution with $20 \mu\text{g ml}^{-1}$ PKA. Glutamate-evoked current responses recorded 2.5 min and 35 min after whole-cell seal formation are not significantly different, showing that PKA did not potentiate the response of mutant GluR6 receptors. *g*, Peak current-voltage relation for GluR6 (S684A) treated with PKA. The peak response to a brief application of $200 \mu\text{M}$ glutamate was recorded at voltages from -100 to $+80$ mV, stepped in 20 mV increments at 1 min intervals. The reversal potential and linearity of the $I-V$ curve is very similar to that seen for the wild-type receptor. *h*, The peak current amplitude in response to rapid perfusion of $200 \mu\text{M}$ glutamate was recorded at 1 min intervals from the two cells shown in *e* and *f* above, and the data was plotted relative to the initial glutamate current response for each cell (PKI ($\bullet$), mutant (∇)).

METHODS. Whole-cell patch clamp recordings²⁶ were made at 25°C from HEK-293 cells expressing GluR6 cDNA. During recordings, cells were continuously superfused with a Krebs solution (control) containing 25 mM HEPES (pH 7.4), 145 mM NaCl, 5.4 mM KCl, 1.8 mM CaCl_2 , 1 mM MgCl_2 , 33 mM glucose. Control to agonist solution exchange was accomplished within $2-3$ ms by solenoid-mediated switching between the two sides of a theta tube positioned within 50 to $100 \mu\text{m}$ of the cell²⁷. In most experiments, cells were lifted from the bottom of the recording chamber immediately after formation of the whole-cell recording mode to optimize rapid perfusion. Patch pipettes ($4-7$ M Ω) used for recording contained 10 mM HEPES (pH 7.1), 140 mM KCl, 1 mM MgCl_2 , 4 mM MgATP, 11 mM EGTA, 1 mM CaCl_2 , 2 mM TEA (control). Recordings were made from HEK-293 cells $48-72$ h after transfection. Owing to variability in receptor expression among different transfections, the peak current amplitude evoked by $200 \mu\text{M}$ glutamate varied from cell to cell. For cells analysed in these experiments, initial (baseline) peak current ranged from $200-2,500$ pA and did not differ significantly between wild-type and mutant GluR6 transfected cells. Voltage-clamp currents were recorded with an Axopatch 200 amplifier (Axon Instruments). Currents were filtered at 2 kHz, sampled at 4 kHz, stored on computer, and both acquired and analysed using PCLAMP software (Axon Instruments). Site-directed mutagenesis was done as described previously²⁸, with the antisense oligonucleotide $5'$ -ACAAGCACAGCTGTCTCTG- $3'$ used to convert Ser 684 into an alanine residue. The fidelity of this substitution was confirmed by DNA sequence analysis.

agonist-evoked current was potentiated by treatment of these neurons with forskolin or cAMP analogues, as well as on direct perfusion of the cells with PKA. In the white perch retinal horizontal cells, application of dopamine, which increases levels of intracellular cAMP, also potentiated kainate-evoked current¹⁷, demonstrating that glutamate receptors can be modulated by other neurotransmitters. The mechanism underlying the potentiation of GluR6 peak current by PKA remains to be determined because the properties of the glutamate-evoked macroscopic current, including the dose-response, $I-V$ relationship, rise time and desensitization kinetics, are unchanged by treatment with PKA. Further study is needed to determine whether phosphorylation of GluR6 by PKA affects channel open time, open frequency or single channel conductance.

The physiological role of ionotropic glutamate receptors composed of GluR6 subunits is not clear. GluR6 messenger RNA is widely distributed in the central nervous system²⁰, with high concentration in the cerebellum and hippocampus, particularly

area CA3. But current responses exhibiting both rapid desensitization and high sensitivity to KA have not yet been recorded from neurons in the brain, although they have been observed in mammalian dorsal root ganglion neurons²³. A possible explanation is that homooligomeric GluR6 receptors in the brain are located on distal dendrites or small interneurons, and thus are inaccessible to recordings from neuronal cell bodies. Alternatively, GluR6 subunits may combine with other glutamate receptor subunits to form ion channels with different physiological properties.

Our data demonstrate that glutamate receptors can be directly phosphorylated and that this phosphorylation regulates their function. It is becoming increasingly clear that protein phosphorylation of ligand-gated ion channels, including nicotinic acetylcholine receptors, GABA_A receptors, glycine receptors and glutamate receptors, is a major mechanism in the regulation of their function and plays a prominent role in the regulation of synaptic plasticity²¹. We have recently learned that similar results have been obtained by Wang *et al.*²⁹. □

Received 16 November 1992; accepted 26 January 1993.

- Collingridge, G. L. & Singer, W. *Trends pharmacol. Sci.* **11**, 290-296 (1990).
- Choi, D. W. *Neuron* **1**, 623-634 (1988).
- Olney, J. W. A. *Rev. Pharmac. Toxicol.* **30**, 47-71 (1990).
- Monaghan, D. T., Bridges, R. J. & Cotman, C. W. A. *Rev. Pharmac. Toxicol.* **29**, 365-402 (1989).
- Gasic, G. P. & Hollmann, M. A. *Rev. Physiol.* **54**, 507-536 (1992).
- Sommer, B. & Seeburg, P. *Trends pharmacol. Sci.* **13**, 291-296 (1992).
- Nakanishi, S. *Science* **258**, 597-603 (1992).
- Kutsuwada, T. *et al. Nature* **358**, 36-41 (1992).
- Linden, D. J. & Connor, J. A. *Science* **254**, 1656-1659 (1991).
- Gerber, G. *et al. J. Neurosci.* **9**, 3606-3617 (1989).
- Chen, L. & Huang, L. M. *Nature* **356**, 521-523 (1992).
- Chen, L. & Huang, L. M. *Neuron* **7**, 319-326 (1991).
- Liman, E. R., Knapp, A. G. & Dowling, J. E. *Brain Res.* **481**, 399-402 (1989).
- Wang, L.-Y., Salter, M. W. & MacDonald, J. F. *Science* **253**, 1132-1135 (1991).
- Greengard, P., Jen, J., Nairn, A. C. & Stevens, C. F. *Science* **253**, 1135-1138 (1991).
- Chavez-Noriega, L. E. & Stevens, C. F. *Brain Res.* **574**, 85-92 (1992).
- Knapp, A. G. & Dowling, J. E. *Nature* **325**, 437-439 (1987).

- Keller, B. U., Hollmann, M., Heinemann, S. & Konnerth, A. *EMBO J.* **11**, 891-896 (1992).
- Madison, D. V., Malenka, R. C. & Nicoll, R. A. *Rev. Neurosci.* **14**, 379-397 (1991).
- Egebjerg, J., Bettler, B., Hermans-Borgmeyer, I. & Heinemann, S. *Nature* **351**, 745-748 (1991).
- Swope, S. L., Moss, S. J., Blackstone, C. D. & Huganir, R. L. *FASEB J.* **6**, 2514-2523 (1992).
- Huganir, R. L., Miles, K. & Greengard, P. *Proc. natn. Acad. Sci. U.S.A.* **81**, 6968-6972 (1984).
- Huettnier, J. E. *Neuron* **5**, 255-266 (1990).
- Bettler, B. *et al. Neuron* **8**, 257-265 (1992).
- Blackstone, C. D. *et al. J. Neurochem.* **58**, 1118-1126 (1992).
- Hamill, O. P. *et al. Pfluegers Arch.* **391**, 85-100 (1981).
- Tang, C.-M., Dichter, M. & Morad, M. *Science* **243**, 1474-1477 (1989).
- Kunkel, T. A., Roberts, J. D. & Zabour, D. L. *Meth. Enzym.* **154**, 367-382 (1987).
- Wang, L.-Y., Taverna, F. A., Huang, X.-P., MacDonald, J. F. & Hampson, D. R. *Science* (in the press).

ACKNOWLEDGEMENTS. We thank S. Heinemann and J. Egebjerg (The Salk Institute) for the cDNA clones of GluR6 and M. D. Uhler for the cDNA clones of the catalytic subunit of the cAMP-dependent protein kinase. Supported by the Howard Hughes Medical Institute (R.L.H.), a CIDA grant from the NIH (L.A.R.) and an MSTP Fellowship (C.D.B.).

Qa-2 molecules are peptide receptors of higher stringency than ordinary class I molecules

Olaf Rötzschke*, Kirsten Falk*, Stefan Stevanović†, Blazenka Grahovac*, Mark J. Soloski‡, Günther Jung† & Hans-Georg Rammensee*§

* Max-Planck-Institut für Biologie, Abteilung Immunogenetik, Corrensstrasse 42, W-7400 Tübingen, Germany

† Institut für Organische Chemie, Universität Tübingen, Auf der Morgenstelle 18, W-7400 Tübingen, Germany

‡ The Johns Hopkins University of Medicine, Division of Molecular and Clinical Rheumatology, 725 North Wolfe Street, 617 Hunterian Building, Baltimore, Maryland 21205, USA

CLASS I molecules of the major histocompatibility complex (MHC) transport peptides to the cell surface for surveillance by T cells¹. Ligand specificity is stringent and differs from allele to allele²⁻⁴. Here we report analysis of natural ligands of 'unconventional' glycoposphatidyl-anchored mouse class I molecules, Qa-2. The function of these molecules is unclear^{5,6}; they can serve as recognition structures for 'unrestricted' cytotoxic T cells but have not been found to present peptides to T cells, although the DNA sequence suggests a similar peptide binding groove to that of 'conventional' class I molecules⁷, and other unconventional class I molecules can present antigens in a few cases⁸⁻¹⁰. Pool sequencing of natural Qa-2 ligands shows that Qa-2 molecules are indeed peptide receptors, having ligand specificity similar to that of conventional class I molecules, that is, a predominant length of nine amino acids, anchor positions, and hydrophobic termination of peptides. But ligand specificity is much more stringent than with other class I molecules: of the nine positions, two are anchors and four have rather limited occupancy.

Spleens of about 400 mice from various strains expressing Qa-2^a were mechanically disrupted and treated with detergent. Qa-2 molecules were immunoprecipitated using solid-phase coupled 1-1-2 antibodies¹¹, and treated with trifluoroacetic acid to dissociate peptide ligands from class I molecules². The soluble part of the resulting material was separated on a reversed-phase high-performance liquid chromatography (HPLC) column (Fig. 1). Fractions 20-28, found earlier to contain peptide ligands of conventional class I molecules², were pooled and sequenced by Edman degradation. Raw data (Table 1a) were analysed as described previously². Very clear amino-acid signals were found at each position from 2 to 9; no or very weak signals at position 10 indicated the majority of Qa-2 ligands to be nonamers. Several positions, such as 7 and 9, contained only one strong signal, or very few strong signals for residues with closely related side chains. Other positions contained some more residues, mostly with very clear signals.

Qa-2 ligands were also isolated from 500 thymi collected from the same mouse strains; sequencing was done from two separate pools, pool one from fraction 20-26 (Table 1b), and pool two from fraction 27-29. Pth-amino acid recoveries were somewhat weaker than with the spleen material but the same sequence pattern determined for splenic peptides was observed, except that the weak signals in the spleen preparation were not detected in the thymus. Pool two yielded essentially the same sequence information as fractions 20 to 29 of lung-derived Qa-2 ligands (data not shown).

A summary of pool sequencing data from these sources is in Table 2. Qa-2 ligands are generally nonamers. Position 7 is a positively charged anchor occupied with His (traces of Arg were found in only one experiment). The carboxy terminus at position

9 is an anchor for hydrophobic residues with long side chains, Leu, Ile and Phe. In addition, a few ligands with different lengths (note occasional weak signals at position 10) cannot be excluded. Positions 2, 3, 5 and 6 are auxiliary anchors (definition in ref. 2), occupied by aliphatic residues, and in addition Gln at 2, Asn at 3, various residues at low frequencies at 5, and a strong Lys at 6. The remaining positions are also preferentially occupied with only a few residues, mostly giving very clear signals. Position 4 has a strong Pro signal, position 8 has exclusively polar residues, mainly Glu, Gln, Asp and Asn, and also in traces Lys, Ser, Arg and Thr. Position 1 is hard to judge with this approach; it appears, however, that Lys is frequently used. Thus, almost every single position of the Qa-2 motif either requires particular residues, or shows preference for certain residues. This indicates that only relatively few peptides fit to this motif, as compared to the enormous number of peptides theoretically fitting to other class I motifs. Thus, one would expect that the number of different natural ligands found on Qa-2 molecules is smaller than that on ordinary class I molecules.

The interpretation of these peptide motif data is complicated by the fact that Qa-2-specific antibodies react with two different gene products on B6 cells, Q7 and Q9 (refs 5, 12). These genes differ from each other at a single amino-acid residue (position 173) that might cause different peptide specificity, although the

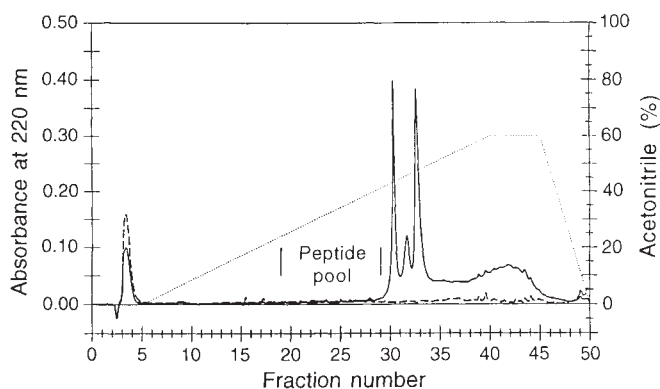


FIG. 1 Reversed-phase HPLC separation of Qa-2 ligands. Qa-2^a ligands of about 400 spleens were separated on a 4.0 mm column. Dotted lines, per cent acetonitrile in the gradient; solid and dashed lines, absorbance at 220 nm; solid lines, Qa-2 eluted material; dashed lines, mock-eluted material. Fractions 20 to 29 (indicated by vertical lines) were pooled for sequencing.

METHODS. Frozen spleens that had been collected from the following strains (all Qa-2^a) were purchased from Harlan Olac Ltd, Blackthorn, UK: C57BL/10ScSn, C57BL/6/J, C57BL/6, C57BL/6bg, C57BL/6-ob, C57BL/6-Tla^a, 129, B10.HTG and BALB.B. The use of this collection of strains rather than a single strain reduced both waiting time and price. Roughly 400 spleens were mechanically disrupted in about 50 ml PBS containing proteinase inhibitors (Aprotinin and Leupeptin (Boehringer-Mannheim), 2 mg ml⁻¹ each; Pepstatin (Boehringer-Mannheim), 0.0002%; PMSF (Sigma), 0.1 mM) using an ultraturax device (IKA) and a motor douncer (B. Braun). This material was brought to 400 ml PBS containing proteinase inhibitors and 1% NP40 (Boehringer-Mannheim) final concentration and stirred for 45 min at 4 °C, then centrifuged (5 min, 800g). Supernatant was spun again (45 min, 180,000g) and filtered (0.45 µm) after floating lipids had been removed. The supernatant was subjected to a column containing glycine-coupled beads for a mock precipitation, then to an affinity column (1 ml CNBr-activated Sepharose (Pharmacia) coupled with 5 mg Qa-2-specific 1-1-2 (ref. 11) antibody) at 0.5 ml min⁻¹ 4 °C. The column was washed as follows. PBS, 0.5% NP40; PBS, 0.1% NP40; threefold bed volume PBS; and again threefold bed volume PBS. Beads were removed from the column, and suspended in 1.5 ml TFA 0.1%, pH was adjusted to 2 by adding 10% TFA (about 10 µl). Beads were shaken for 10 min (23 °C) and spun down in an Eppendorf centrifuge (0.5 min, 14,000 r.p.m.) Supernatant was collected; TFA treatment was repeated twice using 0.75 ml. Combined supernatants were spun again (3 min, 14,000 r.p.m.) to remove residual beads, and were dried in a SpeedVac centrifuge (Savant). HPLC separations were done under the same conditions as described earlier².

§ To whom correspondence should be addressed.

TABLE 1 Sequencing of the ligand mixture eluted from Qa-2 molecules

(a) Qa-2 molecules from spleens*

Cycle	Amino-acid residues (in pmol)																	
	A Ala	R Arg	N Asn	D Asp	E Glu	Q Gln	G Gly	H His	I Ile	L Leu	K Lys	M Met	F Phe	P Pro	S Ser	T Thr	Y Tyr	V Val
1	107.0	4.7	6.9	—	41.0	39.2	64.3	0.0	19.6	14.8	93.0	1.7	10.4	11.2	34.7	24.2	10.6	35.4
2	55.3	4.7	2.4	—	35.5	77.3	45.8	0.0	23.1	82.6	12.6	29.9	8.3	7.5	10.3	0.0	4.1	31.5
3	15.0	3.6	<u>20.6</u>	—	11.3	14.7	30.9	1.2	<u>77.9</u>	<u>132.0</u>	0.7	8.3	4.9	7.6	5.2	<u>14.3</u>	4.9	12.9
4	25.0	3.9	9.5	8.9	32.5	16.8	49.7	0.5	21.7	14.4	7.2	1.3	6.0	<u>22.1</u>	8.6	5.2	3.3	5.5
5	17.2	0.3	5.0	8.1	33.5	22.4	29.6	2.6	68.2	20.9	0.0	2.1	<u>9.7</u>	9.8	5.4	<u>11.3</u>	5.9	69.1
6	14.3	2.9	3.0	6.3	20.0	8.7	16.0	1.3	<u>69.8</u>	<u>36.8</u>	<u>12.7</u>	<u>7.6</u>	<u>10.4</u>	9.0	3.3	7.7	6.7	29.1
7	6.1	<u>4.8</u>	2.0	5.8	7.6	3.2	12.6	25.6	11.3	7.0	0.2	1.6	3.0	4.6	2.2	2.8	2.0	3.2
8	8.4	5.0	<u>13.9</u>	<u>13.4</u>	<u>26.5</u>	<u>11.7</u>	10.6	7.2	6.6	7.3	6.5	2.0	3.1	3.4	<u>5.5</u>	4.4	1.6	4.1
9	4.3	2.0	3.3	6.6	5.8	2.4	10.1	0.9	<u>15.2</u>	<u>53.4</u>	0.0	1.2	<u>11.2</u>	2.9	2.0	1.0	1.5	2.5
10	2.7	1.5	1.7	3.8	2.9	1.7	9.1	1.6	5.3	10.7	0.0	1.1	3.7	2.6	1.5	2.4	0.7	1.3

(b) Qa-2 molecules from thymus†

Cycle	Amino-acid residues (in pmol)																	
	A Ala	R Arg	N Asn	D Asp	E Glu	Q Gln	G Gly	H His	I Ile	L Leu	K Lys	M Met	F Phe	P Pro	S Ser	T Thr	Y Tyr	V Val
1	91.9	2.8	5.2	0.0	32.5	29.2	64.0	0.0	39.6	15.7	24.8	12.9	13.4	14.2	39.2	29.9	17.9	51.4
2	50.9	0.3	2.2	0.0	29.4	44.3	46.5	1.1	22.1	<u>29.2</u>	6.4	<u>24.1</u>	6.8	7.7	16.0	12.5	8.3	28.3
3	22.1	0.7	8.5	2.0	13.6	16.6	39.0	0.7	32.6	<u>53.8</u>	3.1	12.1	7.8	7.5	10.9	12.5	8.1	17.6
4	19.5	3.2	2.7	<u>4.3</u>	15.8	9.2	28.2	0.4	10.6	10.1	1.5	2.9	3.5	<u>11.2</u>	6.1	2.4	3.0	6.9
5	10.6	0.5	1.7	3.9	12.4	7.4	14.8	0.6	20.4	8.7	0.3	1.4	1.8	5.4	4.4	3.3	2.2	<u>19.6</u>
6	11.7	2.4	2.9	3.8	11.4	6.2	12.1	1.2	<u>21.2</u>	8.5	1.1	5.8	1.7	4.8	3.0	2.7	<u>4.5</u>	18.8
7	8.5	0.9	1.7	2.0	7.2	2.6	11.5	<u>4.4</u>	7.5	3.9	0.0	1.8	1.4	3.1	1.9	0.5	4.0	6.3
8‡	9.4	2.0	<u>3.4</u>	<u>4.1</u>	<u>16.0</u>	4.0	12.1	4.0	3.5	4.4	0.2	1.1	1.3	3.4	1.9	0.4	2.0	4.2
9‡	8.3	1.3	<u>3.5</u>	<u>7.5</u>	11.7	3.4	14.9	3.1	<u>7.7</u>	<u>25.7</u>	0.1	0.9	<u>4.9</u>	<u>5.1</u>	0.9	0.2	2.6	7.3
10‡	4.2	<u>3.1</u>	1.5	4.4	5.8	1.3	10.7	2.3	4.3	1.5	0.0	<u>2.1</u>	2.8	2.4	1.6	0.5	2.3	3.5

(c) Q7 gene products from Hepa-Ova-Q7 transfectants§

Cycle	Amino-acid residues (in pmol)																	
	A Ala	R Arg	N Asn	D Asp	E Glu	Q Gln	G Gly	H His	I Ile	L Leu	K Lys	M Met	F Phe	P Pro	S Ser	T Thr	Y Tyr	V Val
1	183.9	1.7	8.7	0.0	71.8	119.1	463.8	5.1	60.3	31.5	36.8	10.6	17.6	14.5	64.8	37.6	22.2	56.9
2	63.4	0.1	3.6	0.0	59.3	116.6	470.0	2.1	48.1	<u>107.0</u>	8.0	<u>39.6</u>	6.8	10.4	21.1	5.8	14.0	35.0
3	25.3	1.4	<u>39.5</u>	<u>13.2</u>	27.1	40.2	317.9	1.2	62.4	<u>121.3</u>	1.9	30.3	5.4	10.9	12.1	<u>10.5</u>	8.1	22.7
4	25.5	1.7	13.6	<u>40.7</u>	44.7	21.0	209.5	1.6	20.2	16.0	5.7	3.6	7.8	63.2	11.6	5.4	6.4	15.2
5	13.1	1.3	7.5	17.6	30.6	25.9	85.0	2.4	<u>50.7</u>	<u>48.8</u>	0.0	2.7	4.5	17.8	4.7	4.0	<u>11.6</u>	<u>68.5</u>
6	15.1	1.0	6.3	21.8	21.6	11.5	47.2	1.2	<u>122.5</u>	44.9	0.0	<u>6.8</u>	<u>12.1</u>	24.2	4.1	5.4	8.3	37.2
7	7.2	2.1	3.6	13.9	8.6	3.3	27.1	<u>35.4</u>	15.6	8.1	0.0	2.2	3.7	12.4	2.5	1.5	3.6	7.1
8	10.4	2.6	<u>24.6</u>	18.1	<u>20.2</u>	<u>10.6</u>	20.1	7.5	8.4	5.2	0.0	3.3	3.0	6.3	<u>5.6</u>	<u>3.2</u>	3.5	7.7
9	5.0	1.3	4.9	7.5	5.5	2.6	15.4	2.6	<u>30.9</u>	<u>32.8</u>	0.0	2.4	<u>11.4</u>	4.4	1.7	2.4	3.1	7.4
10	2.9	0.0	1.7	2.9	1.4	0.9	7.9	2.9	9.3	7.7	0.0	4.7	4.2	3.1	0.9	0.0	3.0	5.7

* Fractions 20 to 29 of HPLC-separated Qa-2 ligands (Fig. 1) were pooled and sequenced by Edman degradation as described². Data analysis was as previously described². Essentially, only increases of 50% or more over the previous pmol value at a given position were considered a significant signal. Arg signals with less than 3.5 pmol were considered as background.

† Qa-2 ligands from 500 thymi of the same source as the spleens (see Fig. 1) were extracted and separated as for the spleen material. Fractions 20 to 26 were pooled and sequenced as above.

‡ Owing to technical problems, cycles 8 and higher are of less confidence; therefore, those residues not also detected in other experiments were not considered for the motif in Table 2.

§ Qa-2 ligands were isolated from about $2-3 \times 10^{10}$ Hepa-Qa-2 cells. These cells had been prepared by transfecting Hepa-1 cells²³ with Q7 cDNA isolated from a λ gt10 cDNA library derived from activated B6 T cells. Q7 cDNA was inserted into the expression vector pcEXO-3 (ref. 24). Qa-2 ligands from Hepa-Ova-Qa-2 cells were extracted in the same way as for spleens, with the following difference: $2-3 \times 10^{10}$ cells were lysed in 400 μ l PBS/detergent without mechanical disruption. The resulting TFA-extracted material was separated as for the splenic material. Fractions 20 to 29 were pooled and sequenced.

substitution of Glu by Gln at 173 (not directly in the binding groove) makes this unlikely. Nevertheless, to account for this possibility, the ligands of the Q7 gene product expressed in transfectant Hepa-Qa-2 cells were also analysed by pool sequencing (Table 1c). The resulting Q7-motif is essentially identical to that of organ-derived Qa-2 molecules, apart from a few minor differences that could be due to minor differences in peptide ligand specificity of Q7 and Q9 gene products or to cell type specific differences in the respective peptide ligands. Thus, it can be concluded that Q7 and Q9 gene products have the same general peptide motif, although some small differences cannot be excluded.

Another limitation of our results might be a bias of the antibody used for precipitating the Qa-2 molecules for particular peptide/Qa-2 combinations, as has been described for an H-2E_αE_β-specific antibody¹³. But we find essentially the same peptide motifs on precipitating Qa-2 molecules with two other antibodies, MA 46 (specific for the α 3-domain; ref. 14) and 1-9-9 (binding to α 1/ α 2 domains; ref. 11).

We conclude that Qa-2 molecules are peptide receptors like

TABLE 2 The Qa-2 specific peptide motif

	Position*								
	1	2	3	4	5	6	7	8	9
Dominant anchor residues							H		L
									I
									F
Strong	K	M	N	P	V	K		E	
		L	I		I	M		Q	
		Q	L			I		N	
								D	
Also detected	A		T	E	L	L	R	K	
	E			A	T	F		S	
	Q			G	E	N		T	
				K	H	Y		R	
				S	M				
				D	F				
					Y				

* Position numbers in bold, anchor; position number underlined, auxiliary anchor (for definition, see ref. 2).

TABLE 3 Protein stretches fitting to Qa-2-motif

(a) Mouse proteins

Origin	Position								
	1	2	3	4	5	6	7	8	9
Complement C9	D	L	N	D	V	K	H	C	L
Retrovirus-related pol polypeptide (reverse transcriptase)	D	L	I	P	I	L	H	K	L
c-ets-2 protein (proto-oncogene)	H	L	N	A	V	P	H	W	I
VDJ recombination-activating protein, rag-1	S	Q	N	L	V	F	H	S	I
Epidermal growth factor precursor	I	Q	N	D	V	G	H	P	F
NADH-ubiquinone oxidoreductase chain 4	T	M	L	M	I	A	H	G	L
Interferon- γ -precursor	E	L	I	R	V	V	H	Q	L
Galactosyl transferase associated protein kinase	V	M	N	Y	V	E	H	D	L

(b) Pathogen-derived proteins

Origin	Position									Organ or tissue affected
	1	2	3	4	5	6	7	8	9	
<i>Schistosoma mansoni</i> haemo- globin precursor	F	L	I	S	I	L	H	I	L	Gut
<i>Schistosoma mansoni</i> 3-hydroxy-3- methyl-glutaryl-coenzyme	N	Q	L	F	V	D	H	M	L	Gut
<i>Salmonella typhimurium</i> DNAB-Protein	D	M	I	A	V	A	H	E	I	Gut
<i>Salmonella typhimurium</i> acetate operon repressor	N	L	L	A	I	V	H	P	I	Gut
<i>Salmonella typhimurium</i> oligopeptide permease protein	S	L	I	F	I	A	H	D	L	Gut
<i>Clostridium tetani</i> tetanus toxin pre- cursor	E	L	I	H	V	L	H	G	L	Muscle*
<i>Klebsiella aerogenes</i> urease alpha subunit	D	M	L	M	V	C	H	H	L	Lung
Human papillomavirus type 1a	R	L	L	L	V	G	H	P	L	Skin
Human papillomavirus type 41	R	L	L	T	V	G	H	P	F	Skin
Human cytomegalovirus	R	Q	I	V	V	V	H	A	L	Epithelia
Vaccinia virus	T	L	I	T	V	R	H	K	L	Skin
Newcastle disease virus	E	L	I	H	V	N	H	L	I	Lung
Bovine viral diarrhoea virus	K	L	L	E	I	F	H	T	I	Gut
Varicella zoster virus strain Dumas	L	L	L	T	I	L	H	H	F	Skin

The Swissprot protein sequence data bank was screened for nonamer peptide stretches having Met, Gln or Leu at 2, Asn, Ile or Leu at 3, Ile or Val at 5, His at 7 and Leu, Ile or Phe at 9. Thus, the auxiliary anchor at position 6, as well as the use of 'weak' residues at position 3 or 5 were not considered, indicating that the search was not completely exhaustive. 131 matches were found this way, 8 of these were from mouse proteins, and 14 from organisms recognized as animal pathogens.

* Reservoir in gut.

other class I molecules. Because Qa-2-expression is affected by the RMA-S mutation (K.F. *et al.*, unpublished results), as is expression of other class I molecules including Hmt and Qa-1 (ref. 15), Qa-2 expression must be influenced by TAP gene products¹⁶. Thus, the mechanisms of antigen processing for Qa-2 and other class I molecules are probably similar. The interactions of peptides and Qa-2 molecules seem essentially similar to those of other class I molecules. The N and C termini as well as the C-terminal side chain are probably held in the same sites as for B27, A2, K^b or other molecules¹⁷⁻¹⁹. Interaction of the other conserved side chains with Qa-2 sites cannot be easily predicted; it is likely, however, that the His side chain at position 7 is accommodated by a specific pocket close to the C terminus. Because no obvious negative charge stands out in the Q7 or Q9 sequences at that area, one might speculate that Ser at 73, Ser at 77 and Trp at 97 might form a pocket accommodating His side chains through hydrogen bonds and aromatic interactions. An answer to this must await analysis of a Qa-2 crystal.

The stringent peptide ligand specificity of Qa-2 molecules,

and the likelihood that only a few different peptides are presented by these molecules, indicates that there must be a highly specialized function for Qa-2 molecules. Another unconventional class I molecule, Hmt (now H-2M3), is highly specialized in presenting N-formylated peptides of procaryotic or mitochondrial origin²⁰⁻²². Screening of the Swissprot data bank for peptide stretches fitting to the two Qa-2 anchors, His at 7 and Leu, Ile or Phe at 9, resulted in 1,789 matches in mouse proteins. This compares to 8,414 matches if the same is done for HLA-A2 anchors (Leu or Met at 2 and Val or Leu at 9) and reflects the relatively rare use of His residues. Including the respective auxiliary anchors into the search, the different frequency of fitting peptide stretches (again from mouse proteins) is even more striking: 2,086 matches for the A2 motif (Val, Ile, Met or Leu at position 67, in addition to above anchors) versus only 8 matches (Table 3a) for the Qa-2 motif (with Met, Glu or Leu at 2, Asn, Ile or Leu at 3, Ile or Val at 5, His at 7 and Leu, Ile or Phe at 9, that is, only 3 of the 4 auxiliary anchors considered). Thus, one of the Qa-2 functions might be to present a small set of self peptides. The same Qa-2 motif parameters were used to screen the entire data base, consisting of 26,000 gene entries that code for 9 million amino acids. Only 131 matches were found in all; screening with the A2 motif (not done) would result in an estimated 32,000 matches. This comparison again underlines the stringent peptide specificity of Qa-2 molecules. Of the 131 matches, 14 were recognized as belonging to pathogens (Table 3b). All of these nonapeptides are from pathogens that affect epithelia and/or are gut-associated. Because $\gamma\delta$ T cells of limited diversity locate to several types of epithelia²⁵ and because unconventional class I molecules (Tla) are expressed in intestinal epithelium^{26,27}, it has been speculated that $\gamma\delta$ T cells might be specialized to recognize peptides from these tissues on unconventional class I molecules^{26,27}. Our data bank search results are now compatible with this notion and specify it to the hypothesis that Qa-2 molecules might be specialized to present epithelial-pathogen derived antigens to $\gamma\delta$ T cells. □

Received 11 September; accepted 1 December 1992.

- Röttschke, O. & Falk, K. *Immun. Today* **12**, 447-455 (1991).
- Falk, K., Röttschke, O., Stevanovic, S., Jung, G. & Rammensee, H.-G. *Nature* **351**, 290-296 (1991).
- Jardetzky, T. S., Lane, W. S., Robinson, R. A., Madden, D. R. & Wiley, D. C. *Nature* **353**, 326-329 (1991).
- Hunt, D. F. *e. Science* **255**, 1261-1263 (1992).
- Flaherty, L., Elliott, E., Tine, J. A., Walsh, A. C. & Waters, J. B. *CRC Crit. rev. Immun.* **10**, 131-175 (1990).
- Mellor, A. L. *et al. Int. Immun.* **3**, 493-502 (1991).
- Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L. & Wiley, D. C. *Nature* **329**, 506-512 (1987).
- Vidovic, D., Roglic, M., McKune, K., Guerdner, S., MacKay, C. & Dembic, Z. *Nature* **340**, 646-650 (1989).
- Loveland, B., Wang, C. R., Yonekawa, H., Hermel, E. & Lindahl, K. F. *Cell* **60**, 971-980 (1990).
- Milligan, G. N., Flaherty, L., Braciale, V. L. & Braciale, T. J. *J. exp. Med.* **174**, 133-138 (1991).
- Sharrow, S. O., Arn, J. S., Stroyanowski, I., Hood, L. & Sachs, D. H. *J. Immunol.* **142**, 3495-3502 (1989).
- Soloski, M. J., Hood, L. & Stroyanowski, I. *Proc. natn. Acad. Sci. U.S.A.* **85**, 3100-3104 (1988).
- Rudensky, A. Y., Preston-Hurlburt, P., Hong, S. C., Barlow, A. & Janeway, C. A. *Nature* **353**, 622-627 (1991).
- Hasenkamp, K. J., Cory, J. M. & Stimpfling, J. H. *Immunogenetics* **25**, 136-139 (1987).
- Hermel, E., Grigorenko, E. & Lindahl, K. F. *Ing. Immun.* **3**, 407-412 (1991).
- Monaco, J. J. *Immun. Today* **13**, 173-179 (1991).
- Madden, D. R., Gorga, J. C., Strominger, J. L. & Wiley, D. C. *Nature* **353**, 321-325 (1991).
- Saper, M. A., Bjorkman, P. J. & Wiley, D. C. *J. molec. Biol.* **219**, 277-319 (1991).
- Fremont, D. H., Matsamura, M., Stura, E. A., Peterson, P. A. & Wilson, I. A. *Science* **257**, 919-927 (1992).
- Shawar, S. M., Vyas, J. M., Rodgers, J. R., Cook, R. G. & Rich, R. R. *J. exp. Med.* **174**, 941-944 (1991).
- Kurlander, R. J., Shawar, S. M., Brown, M. L. & Rich, R. R. *Science* **257**, 678-679 (1992).
- Pamer, E. G., Wang, C.-R., Flaherty, L., Fischer, Lindahl, K. & Bevan, M. J. *Cell* **70**, 215-223 (1992).
- Darlington, G., Bernhard, H., Miller, R. & Ruddle, F. J. *natn. Cancer Inst.* **64**, 809-819 (1990).
- Miller, J. & Germain, R. J. *exp. Med.* **164**, 1478-1489 (1986).
- Allison, J. P. & Havran, W. L. *A. Rev. Immun.* **9**, 679-705 (1991).
- Hershberg, R. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 9727-9731 (1990).
- Wu, M., Van Kaer, L., Itohara, S. & Tonegawa, S. *J. exp. Med.* **174**, 213-218 (1991).

ACKNOWLEDGEMENTS. We thank J. Klein for support, J. Venachio for isolation of the Q7 DNA, D. Sachs, G. Waneck and K. Udaoka for the 1-1-2 and 1-9-9 hybridomas, H.-J. Bühring for cell sorting and S. Faath for technical assistance. This work was supported by a grant from the NIH (to M.J.S.), by the Deutsche Forschungsgemeinschaft (Sonderforschungsbereiche 120 and 323) and by a grant from Bundesministerium für Forschung und Technologie (to H.-G.R.).

Drug assay using antibody mimics made by molecular imprinting

George Vlatakis*, Lars I. Andersson, Ralf Müller* & Klaus Mosbach†

Department of Pure and Applied Biochemistry, Chemical Centre, University of Lund, PO Box 124, S-221 00 Lund, Sweden

LIGAND-BINDING assays are used for determination of minute amounts of substances in the bloodstream. Such assays require a receptor that specifically binds the substance of interest. The receptor used is often an antibody¹⁻⁵, but antibodies require special handling and a costly production procedure⁵. We have used molecular imprinting, a method for creating selective recognition sites in synthetic polymers⁶⁻⁸, to prepare polymers that mimic antibody combining sites. Molecular imprints made against theophylline⁹ and diazepam¹⁰ showed strong binding and cross-reactivity profiles similar to those of antibodies. Here we describe a new radiolabelled ligand-binding assay, the molecularly imprinted sorbent assay, which uses antibody mimics. This assay accurately measures drug levels in human serum, with results comparable to those obtained using a well established immunoassay technique. Antibody mimics, which are stable and readily prepared by molecular imprinting, may provide a useful general alternative to antibodies.

Molecular imprinting involves polymerization of functional monomers in the presence of a 'print' molecule (Fig. 1). Removal of the molecule from the rigid polymer results in sites within the polymer that are complementary to, and have an affinity for, the original print molecule. Molecular imprinting has practical applications in enantiomeric separations¹¹⁻¹⁵, especially in the resolution of racemic drugs such as β -blockers¹⁶. We wanted to improve affinity and selectivity of imprinted polymers until they were comparable to biological antibodies. We have used such polymers as substitutes for antibodies in an immunoassay-like technique. We chose two chemically unrelated drugs, theophylline (a bronchodilator)⁹ and diazepam (a tranquillizer, for example valium)¹⁰, as print molecules. The polymers were prepared using methacrylic acid (MAA) as the functional monomer and ethylene glycol dimethacrylate (EDMA) as the crosslinking monomer (Fig. 1). The carboxylic acid function of MAA forms ionic interactions with amino groups¹² and hydrogen bonds with polar functions of the print molecule¹⁴. Dipole-dipole and hydrophobic interactions may also contribute.

The solvent compositions giving optimal binding and selectivity were determined for each polymer (Fig. 2). In general^{14,17}, the substrate binds more strongly to the polymer in a more apolar solvent, and small amounts of acetic acid added to the solvent suppress nonspecific binding. The equilibrium dissociation constants (K_D) for binding of the drugs to the corresponding polymers were estimated by Scatchard plot analysis of binding data for radiolabelled ligands. In both cases, nonlinear plots were obtained because of multiple K_D values, varying from high to low affinity (Fig. 2). We believe that, like polyclonal antibodies, the polymers contain a heterogeneous population of sites with different affinities for the print molecule.

Polymers prepared against theophylline or diazepam were used as antibody substitutes to make competitive binding assays for theophylline and diazepam determination in human serum. The method relies on the inhibition of radiolabelled ligand binding by the serum analyte. The amount of labelled ligand bound to the polymer is inversely related to the concentration of drug present in the sample. Drug-free serum samples spiked

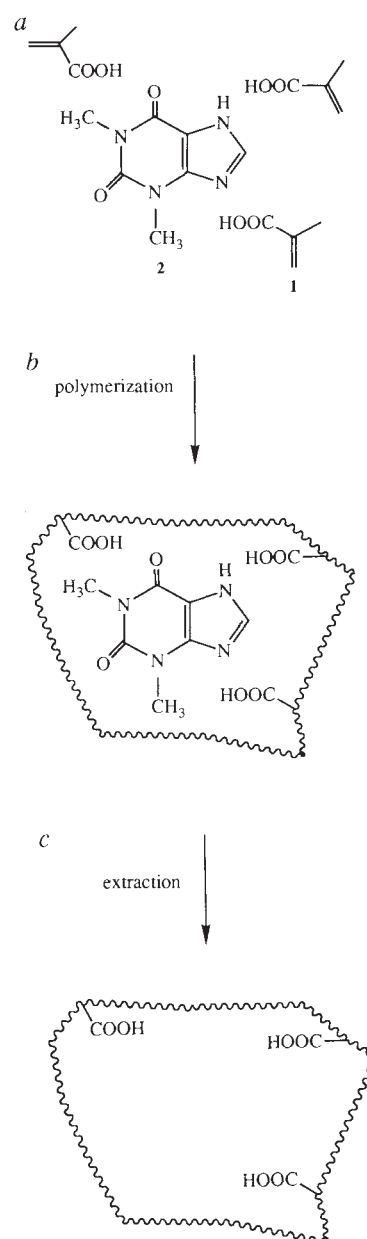


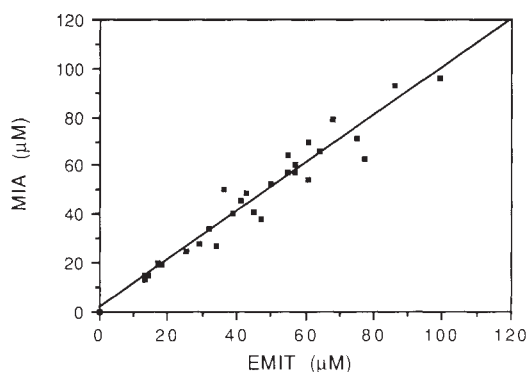
FIG. 1 Schematic representation of the preparation of molecularly imprinted polymers. *a*, Functional monomer MAA (1), is mixed with print molecule, here theophylline (2), and EDMA, the crosslinking monomer, in a suitable solvent. MAA is selected for its ability to form hydrogen bonds with a variety of chemical functionalities of the print molecule. *b*, The polymerization reaction is started by addition of initiator (2,2'-azobis (2-methylpropionitrile), AIBN). A rigid insoluble polymer is formed. 'Imprints', which are complementary to the print molecule in both shape and chemical functionality, are now present within the polymeric network. *c*, The print molecule is removed by solvent extraction. The wavy line represents an idealized polymer structure but does not take into account the accessibility of the substrate to the recognition site. **METHODS.** Anti-theophylline polymer. To a glass bottle were added chloroform (250 ml), theophylline (4.7 g), MAA (9 g), EDMA (93.5 g) and AIBN (1.2 g). The mixture was degassed under vacuum in a sonicating water-bath and sparged with nitrogen for 5 min before polymerization at 60 °C for 24 h. The bulk polymer was ground in a mechanical mortar and wet-sieved (water) through a 25- μ m sieve. The fines were removed by repeated sedimentation in acetonitrile. The print molecule (theophylline) was extracted by extensive washing of the particles with methanol/acetic acid (9/1, v/v). Finally, the polymer particles were dried under vacuum and stored in a desiccator. Anti-diazepam polymer. Diazepam (1.27 g) was mixed with MAA (2.26 g), EDMA (26.1 g) and AIBN (0.5 g) in chloroform (39 ml). The method was as described, except that ultraviolet-initiated (366 nm) polymerization at 4 °C for 16 h was used.

* Present addresses: Foundation for Research and Technology, Institute of Molecular Biology and Biotechnology, 711 10 Heraklion, Crete, Greece (G.V.); Johnson & Johnson Medical, Ostrasse 1, 2000 Norderstedt, Germany (R.M.).

† To whom correspondence should be addressed.

FIG. 2 Comparison of the enzyme-multiplied immunoassay technique (EMIT)²⁶ and MIA competitive binding assays for determination of serum concentration of theophylline in patient samples ($n=32$). Slope, 0.99; intercept, 1.50 μM ; correlation coefficient, 0.98.

METHODS. All EMIT assays were done according to the method of the manufacturer. Assay conditions for MIA were established by applying protocols similar to those used for the optimization of immunoassays using antibodies³⁰. Sample (40 μl) mixed with 40 μl HCl (0.2 M) was extracted with 1 ml dichloromethane-isopropanol (4/1, v/v). The organic layer was evaporated at 40 °C under a stream of nitrogen and redissolved in 100 μl of acetonitrile/acetic acid (99/1, v/v) containing [³H]theophylline (5 ng, 18.6 Ci mmol⁻¹). Anti-theophylline polymer was added (12.5 mg of polymer in 0.9 ml of the same solvent), and the mixture was incubated for 15 h at room temperature. The binding equilibrium was reached after 8 h; 80% of binding occurred within 3 h and 90% within 5 h. After centrifugation, the unbound [³H]theophylline in 200 μl of the supernatant was measured by liquid scintillation counting. The calibration graph was linear over the range 14–224 μM (correlation coefficient, 0.999) with a detection limit of 3.5 μM . The diazepam assay, performed in a similar manner using 5 mg polymer in toluene-heptane (3/1; v/v), was linear from 0.44 to 28 μM (correlation coefficient=0.991) with a detection limit of 0.2 μM . Scatchard plots were analysed by nonlinear least-square fitting with multisite models using the EBDA and LIGAND programmes (Elsevier-Biosoft). For theophylline: $K_D=0.35\pm0.048$ μM and 65 ± 16 μM (associated with a population of sites of



0.016 $\mu\text{mol g}^{-1}$ and 1.3 $\mu\text{mol g}^{-1}$, respectively). For diazepam: $K_D=18\pm7.3$ nM, 2.3 ± 1.3 μM and 60 ± 74 μM (6.2 ± 2.4 nmol g⁻¹, 0.17 ± 0.11 and 1.2 ± 0.93 $\mu\text{mol g}^{-1}$). A three-site model gave better agreement with experimental data (but larger errors in the estimated K_D values) than did a two-site model in the diazepam case, but could not be applied to the theophylline binding data.

with known amounts of theophylline or diazepam were used to establish standard calibration curves. Before the actual assay, the drug was extracted from the serum following standard protocols used for HPLC analysis of these compounds^{18–20} (Fig. 2). The molecularly imprinted sorbent assay (MIA) for theophylline was linear over the range 14–224 μM , which is satisfactory for therapeutic monitoring of the drug^{9,21–23}. The results for diazepam were linear over the range normally used in standard immunoassay techniques for benzodiazepines (0.44–28 μM)^{24,25}. The specificity of the method was tested by the determination of cross-reactivity of major metabolites and of drugs structurally related to theophylline or diazepam (Table 1). The MIA method for theophylline (1,3-dimethylxanthine) seems to be highly specific: of all the compounds tested only 3-methylxanthine showed cross-reactivity. In the case of the diazepam assay several other benzodiazepines showed significant cross-reactivity. This was expected, however, because benzodiazepines have very similar structures (Fig. 3), and even

antibodies have difficulty distinguishing between them^{24,25} (Table 1).

The accuracy of MIA measurement of theophylline was evaluated by analysing 32 patient serum samples. The samples were also analysed with the enzyme-multiplied immunoassay technique (EMIT)²⁶; comparison of the results showed excellent correlation between the two methods (Fig. 2). Furthermore, the reliability of the assay was determined by measurement of theophylline samples of known concentration (three clinically significant concentrations, eleven repetitions, coefficient of variation $\leq 6.5\%$). The MIA method is more cumbersome than the EMIT method as it includes an extraction step (as does HPLC

TABLE 1 Cross-reactivity of xanthine and uric acid derivatives for binding of [³H]-theophylline and benzodiazepines for binding of [³H]-diazepam to imprinted polymers

Theophylline antibodies		Cross-reaction (%)	
Competitive ligand		Abm	Ab*
Theophylline (1,3-dimethylxanthine)		100	100
3-Methylxanthine		7	2
Xanthine		<1	<1
Hypoxanthine		<1	<1
7-(β -Hydroxyethyl)-1,3-dimethylxanthine		<1	<1
Caffeine (1,3,7-trimethylxanthine)		<1	<1
Theobromine (3,7-dimethylxanthine)		<1	<1
Uric acid		<1	<1
1-Methyluric acid		<1	<1
1,3-Dimethyluric acid		<1	<1
Diazepam antibodies		Cross-reaction (%)	
Competitive ligand		Abm	Ab†
Diazepam		100	100
Alprazolam		40	44
Desmethyldiazepam		27	32
Clonazepam		9	5
Lorazepam		4	1
Chlordiazepoxide		2	<1

Ligands were added to drug-free serum and assayed as described for Fig. 2. Cross-reactivities are expressed as the molar ratio (of theophylline and diazepam, respectively, to ligand) giving 50% inhibition of radiolabelled ligand binding. Abm, antibody-combining site mimic; Ab, natural antibody.

* Data from ref. 23.

† Data from ref. 25.

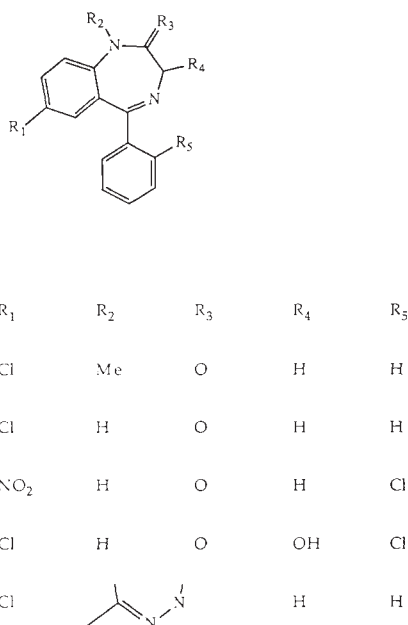


FIG. 3 Structures of the benzodiazepine derivatives.

analysis¹⁸⁻²⁰), incubation and handling of radioactive materials (as do radioimmunoassays). On the other hand, one advantage of molecularly imprinted polymers is their simple and rapid (2-3 days) preparation and their remarkable stability. They can be stored in the dry-state at ambient temperatures for several years without loss of recognition capabilities¹⁷. In addition, the potential to reuse the polymers may be valuable.

The results presented here demonstrate, for the first time, the use of chemically prepared macromolecules with preselected specificity, instead of the traditional biomolecules, as receptors in competitive binding assays. Apart from the practical importance of the described preparations, structural studies on the interactions of drugs with their artificial receptors could yield

valuable insight into the nature of the molecular recognition phenomena²⁷⁻²⁹. Molecular imprints may be made against a great number of organic molecules, for example, drugs, metabolites, hormones and toxins, provided the compound in question has suitable potential interaction points^{14,17}. Preparation of molecular imprints against proteins, DNAs and other biomacromolecules is under investigation. It is noteworthy that molecularly imprinted polymers provide a potential alternative to the use of laboratory animals for the production of antibodies. Preliminary data from studies with an emphasis on recognition in aqueous systems, using compounds such as opiates and biologically active peptides, indicate that this technique may have many applications. □

Received 1 September; accepted 3 December 1992.

1. Yalow, R. S. & Berson, S. A. *Nature* **194**, 1648-1649 (1959).
2. Köhler, G. & Milstein, C. *Nature* **256**, 495-497 (1975).
3. Oellerich, M. *J. clin. Chem. Clin. Biochem.* **22**, 895-904 (1984).
4. Gosling, J. P. *Clin. Chem.* **36**, 1408-1427 (1990).
5. Kurstak, E. in *Enzyme Immunodiagnosis* (ed. Kurstak, E.) 5-11 (Academic, London, 1986).
6. Sellergren, B., Ekberg, B. & Mosbach, K. *J. Chromatogr.* **347**, 1-10 (1985).
7. Wulff, G. *Am. chem. Soc. Symp. Series* **308**, 186-230 (1986).
8. Shea, K. J. & Sasaki, D. Y. *J. Am. chem. Soc.* **113**, 4109-4120 (1991).
9. Hendeles, L., Weinberger, M. & Johnson, G. *Clin. Pharmacokinet.* **3**, 294-312 (1978).
10. Harvey, S. L. in *The Pharmacological Basis of Therapeutics* (eds Gilman, A. G., Goodman, L. S., Rall, T. W. & Murad, F.) 339-351 (Dekker, New York, 1985).
11. Ekberg, B. & Mosbach, K. *Trends Biotechnol.* **7**, 92-96 (1989).
12. Sellergren, B., Lepistö, M. & Mosbach, K. *J. Am. chem. Soc.* **110**, 5853-5860 (1988).
13. O'Shannessy, D. J., Ekberg, B., Andersson, L. I. & Mosbach, K. *J. Chromatogr.* **470**, 391-399 (1989).
14. Andersson, L. I. & Mosbach, K. *J. Chromatogr.* **516**, 313-322 (1990).
15. Wulff, G. & Minarik, M. *J. liq. Chromatogr.* **13**, 2987-3000 (1990).
16. Fischer, L., Müller, R., Ekberg, B. & Mosbach, K. *J. Am. chem. Soc.* **113**, 9358-9360 (1991).
17. Andersson, L. I. thesis, Lund Univ. (1991).

18. Meffin, P. J. & Miners, J. O. in *Progress in Drug Metabolism* Vol. 4 (eds Bridges, J. W. & Chasseaud, L. F.) 261-307 (Wiley, London, 1980).
19. Peng, G. W., Gadalla, M. A. F. & Chiou, W. L. *Clin. Chem.* **24**, 357-361 (1978).
20. Mura, P., Pirou, A., Frailion, P., Papet, Y. & Reiss, D. *J. Chromatogr.* **416**, 303-310 (1987).
21. Castro, A., Ibanez, J., Voight, W., Noto, T. & Malkus, H. *Clin. Chem.* **24**, 944-946 (1978).
22. Chang, J., Gotcher, S. & Gushaw, J. B. *Clin. Chem.* **28**, 361-367 (1982).
23. Poncelet, S. M. et al. *J. Immunoassay*, **11**, 77-88 (1990).
24. Baselt, R. C. in *Advances in Analytical Toxicology* Vol. 1 (ed. Baselt, R. C.) 81-123 (Biomedical, Foster City, CA, 1984).
25. Altunkaya, D. & Smith, R. N. *Forensic Sci. Int.* **39**, 23-37 (1988).
26. Dietzler, D. N. et al. *Clin. chim. Acta* **101**, 163-181 (1980).
27. Cram, D. J. *Nature* **356**, 29-36 (1992).
28. Rebek, J. Jr. *Angew. chem. int. Ed. engl.* **29**, 245-255 (1990).
29. Deslongchamps, G., Galán, A., de Mendoza, J. & Rebek, J. Jr. *Angew. chem. int. Ed. Engl.* **31**, 61-63 (1992).
30. Tijssen, P. *Laboratory Techniques in Biochemistry and Molecular Biology, Practice and Theory of Enzyme Immunoassays* 5th printing (Elsevier, Amsterdam, 1988).

ACKNOWLEDGEMENTS. We thank L.-G. Nilsson, P. Höglund and A. Lidén for serum samples and discussion, and I. A. Nicholls for help in the preparation of the manuscript.

Efficient adenovirus-mediated transfer of a human minidystrophin gene to skeletal muscle of *mdx* mice

Thierry Ragot*, Nathalie Vincent†, Philippe Chafey†, Emmanuelle Vigne*, Hélène Gilgenkrantz†, Dominique Couton‡, Jean Cartaud§, Pascale Briand‡, Jean-Claude Kaplan†, Michel Perricaudet* & Axel Kahn†

*URA 1301 CNRS, Institut Gustave Roussy, PR2, 39 rue Camille Desmoulins, 94805 Villejuif, France

†U 129 INSERM, Institut Cochin de Génétique Moléculaire (ICGM), 24 rue du faubourg St-Jacques, 75014 Paris, France

‡CJF 9003 INSERM, ICGM, 22 rue Méchain, 75014 Paris, France

§UMR 3 CNRS, Institut Jacques Monod, 2 Place Jussieu, 75251 Paris Cedex 05, France

DUCHENNE progressive muscular dystrophy is a lethal and common X-linked genetic disease¹ caused by the absence of dystrophin^{2,3}, a 427K protein encoded by a 14 kilobase transcript⁴. Two approaches have been proposed to correct the dystrophin deficiency in muscle. The first, myoblast transfer therapy, uses cells from normal donors⁵⁻⁷, whereas the second involves direct intramuscular injection of recombinant plasmids expressing dystrophin⁸. Adenovirus is an efficient vector for *in vivo* expression of various foreign genes⁹⁻¹³. It has recently been demonstrated that a recombinant adenovirus expressing the lac-Z reporter gene can infect stably many mouse tissues, particularly muscle and heart^{12,13}. We have tested the ability of a recombinant adenovirus, containing a 6.3 kilobase pair Becker-like dystrophin complementary DNA¹⁴ driven by the Rous sarcoma virus promoter to direct the expression of a 'minidystrophin' in infected 293 cells and C2

myoblasts, and in the *mdx* mouse^{15,16}, after intramuscular injection. We report here that *in vivo*, we have obtained a sarcolemmal immunostaining in up to 50% of fibres of the injected muscle.

X-linked Duchenne progressive muscular dystrophy (DMD) affects 1 in 3,500 male newborns. In an allelic form of DMD, Becker muscular dystrophy (BMD), dystrophin is present, although qualitatively and/or quantitatively abnormal¹⁷, resulting in a less severe phenotype. We used a 6.3 kilobase pair (kbp) human dystrophin cDNA (provided by K. Davies) previously described in a Becker patient with very mild clinical features¹⁴. This cDNA is deleted for exons 17 to 48 (nucleotide (nt) 2,201 to nt 7,306) and encodes a 'minidystrophin' (220K) lacking most of the central repetitive domain of the entire dystrophin (Fig. 1a). We cloned this cDNA under the control of the Rous sarcoma virus (RSV) promoter in a viral vector based on adenovirus (Ad) serotype 5. It is possible to substitute up to 7.5 kbp of exogenous DNA into the Ad genome, as the result of the deletion of the E1 and part of the E3 regions¹⁸. The recombinant Ad bearing this cDNA, Ad-RSVmDys, was obtained by *in vivo* recombination as previously described^{13,19,20}. It is replication-defective in human cells, thus virus stocks were prepared in 293 cells, a human cell line which complements the E1 defect²¹.

These cells were infected with either the recombinant Ad-RSVmDys or the recombinant Ad-RSVβgal (ref. 13; expressing a nuclear-targeted β-galactosidase) as a negative control. A protein of the expected molecular mass (around 200K) was detected by immunotransfer from extracts of Ad-RSVmDys-infected 293 cells but not from extracts of Ad-RSVβgal-infected 293 cells, thus confirming the integrity of the minidystrophin gene construct (Fig. 2a). The expression of the foreign genes *in vivo* is not restricted to cells complementing the E1-defect, but can also be obtained in various cell lines infected with this type of defective Ad^{20,22}. Therefore we examined the ability of the recombinant Ad-RSVmDys to produce the minidystrophin after infection of C2 mouse myoblasts. A signal of the expected size was again detected on immunoblots (Fig. 2b).

TABLE 1 Percentage of dystrophin-positive myofibres in injected *mdx* muscle

Animal number	Age at injection (days postnatal)	Age at death (weeks postnatal)	Injected virus dose* ($\times 10^9$ p.f.u.)	Dystrophin-positive fibres†		
				injected muscle	contralateral	
				<i>T. marm.</i> antibody	P20 AB antibody	<i>T. marm.</i> antibody
36	5	3	5 (a)	50%	<1%	<1%
85	7	4	2.5 (b)	10%	<1%	<1%
88	7	4	2.5 (b)	6%	<1%	<1%
94	7	5	2.5 (b)	5%	<1%	<1%
93	9	2.5	3.3 (c)	30%	<1%	<1%
111	7	2	1.9 (d)	6%	<1%	<1%
115	7	3	1.9 (d)	9%	<1%	<1%
121	7	3	3 (e)	33%	<1%	<1%
125	7	5	3 (e)	24%	<1%	<1%
138	7	14	3 (e)	20%	<1%	<1%

* Five independent stocks of Ad-RSVmDys, titrated by plaque assay on 293 cells, were used as indicated by the letters in brackets.

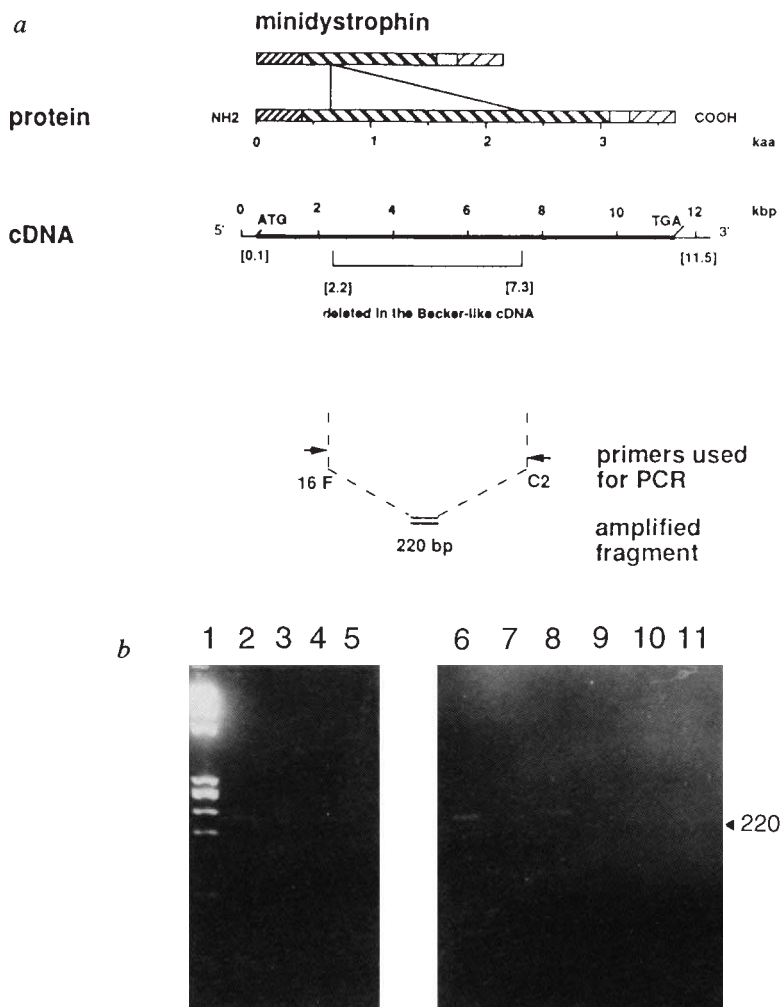
† Percentage of dystrophin-positive myofibres detected by immunofluorescence with the antibody raised against dystrophin of *T. marmorata*, recognizing both the minidystrophin and the entire dystrophin, and P20 AB, recognizing only the entire dystrophin. It was established by scoring around 1,000 fibres in 5 fields (magnification 25) with 150–200 fibres per field.

Because the minidystrophin was produced in myoblasts under *ex vivo* conditions, we tested the ability of the recombinant Ad-RSVmDys to direct *in vivo* expression of this protein after intramuscular injection in *mdx* mice. Animals were injected once 5 to 9 days after birth in the biceps femoris and killed at an age of 2 to 14 weeks. Data for all tested animals are presented in Table 1. The expression of the minidystrophin gene was first assessed in the injected muscle by reverse transcriptase polymerase chain reaction (RT-PCR) with DNase I-treated total

RNAs as a template. Primers were specific to sequences flanking the deleted region in the 6.3 kbp cDNA (Fig. 1a). The expected 220 bp fragment was amplified from the Ad-RSVmDys-injected muscle material (Fig. 1b). In contrast, no signal was found when the reverse transcriptase step was omitted, thus demonstrating that the signal did not arise from contaminating DNA. The signal was also generally undetectable in the contralateral, non-injected muscle. This is in line with the occasional presence (about 100- to 1,000-fold less than in the injected muscle) of

FIG. 1 a, Schematic representation of the human minidystrophin cDNA and protein and localization of the C2 and 16F oligonucleotides used in the PCR experiments. 16F (5'-GCA-TGGCTGGATAACTTTGCCCGG-3') is identical and C2 (5'-TAAACGGTTTACCGCCTTCCACTCAGA-3') is complementary to the dystrophin sequence in exon 16 and 50, respectively. b, Specific RT-PCR amplification of the minidystrophin transcript from injected *mdx* mouse muscle. Lane 1, Molecular mass marker: *Hae*III-restricted $\phi \times 174$ (Promega). Injected muscle (lanes 2, 3, 6, 7, 8, 9), contralateral non-injected muscle (lanes 4, 5), *mdx* muscle injected with Ad-RSV β gal vector (lanes 10, 11). PCR was done with (lanes 2, 4, 6, 8, 10) or without (lanes 3, 5, 7, 9, 11) a reverse transcriptase step.

METHODS. Ad preparations to be inoculated were first purified by separating twice in a CsCl density gradient to avoid carrying over impurities from the cell culture and were extensively dialysed and diluted in PBS before injection to the animals. 5-to-9-day-old *mdx* mice were injected in the biceps femoris with 50 μ l of the viral suspension. Five independently prepared viral stocks were used in these experiments (virus concentrations ranging from 0.4 to 1×10^{11} p.f.u. ml^{-1}). Animals were killed 1 to 13 weeks after injection. RNAs from isolated muscle were extracted by RNAzol B (Bioprobe) according to the manufacturer's instructions. 500 ng of total RNAs were treated by 10 units of DNase I (Boehringer), and were precipitated twice. Pellets were resuspended in 19 μ l PCR buffer ($1 \times = 16.6$ mM NH_4HSO_4 , 50 mM β -mercaptoethanol, 67 mM Tris-HCl (pH 8.8), 6.7 mM EDTA, 4.5 mM MgCl_2), 100 pM random hexanucleotides, 20 units of RNasin (Promega) and 200 μ M each dNTP. In the RT-plus samples, 200 units of Moloney's virus reverse transcriptase (BRL) were added, and cDNA synthesis was done for 1 h at 42 °C. PCR, 35 cycles (95 °C for 30 s, 55 °C for 30 s, 72 °C for 1 min) in 50 μ l containing 20 μ l DNase I-treated RNA sample (with or without cDNA synthesis), PCR buffer $1 \times$, 200 μ M each dNTP, 1 unit of AmpliTaq DNA polymerase (Cetus) and 100 ng of each specific primers.



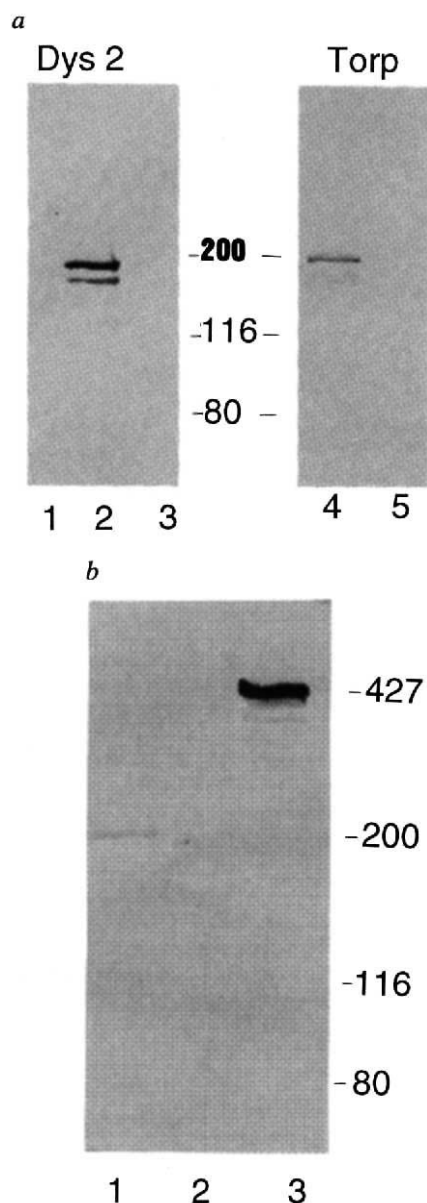


FIG. 2 Immunodetection of the minidystrophin in extracts of *ex vivo* infected cells. *a*, 293 cells 24 h after infection with Ad-RSVmDys. The expression of the minidystrophin cDNA can be detected by immuno-western blot (lanes 2 and 4). We used cells infected with Ad-RSV β gal as a negative control (lanes 1 and 5). Molecular mass markers, lane 3. *b*, C2 myoblasts 60 h after infection with Ad-RSVmDys (lane 1). Negative control, C2 myoblasts infected by Ad-RSV β gal (lane 2). Normal mouse muscle, lane 3. METHODS. 0.2 to 1×10^6 293 cells were broken by three cycles of thawing-freezing and homogenized in $550 \mu\text{l}$ of 10 mM Tris-HCl (pH 6.8) and 1 mM iodoacetamide. The protein concentration was estimated by the method of Lowry and adjusted to 1 mg ml^{-1} . Finally, 1 volume of sample treatment buffer was added. For C2 cells, the presence of myosin comigrating with the minidystrophin necessitates a prior microsome preparation. Cells were homogenized in $500 \mu\text{l}$ 100 mM Tris-HCl (pH 7.45), 10 mM EDTA, 1 mM iodoacetamide, 0.25% Triton X-100, 0.25 M sucrose, then centrifuged for 30 min at $14,000 \text{ r.p.m.}$ in rotor type 70.1 Ti on a Beckman TL 100. The supernatant was removed for a further centrifugation. The pellet was resuspended in $500 \mu\text{l}$ of the same buffer and centrifuged again. The supernatant was removed, pooled with the first one and centrifuged for 60 min at $60,000 \text{ r.p.m.}$ The pellet was finally resuspended in $100 \mu\text{l}$ 10 mM Tris-HCl (pH 6.8) and 1 mM iodoacetamide. The samples were then treated as previously described²⁸. We used a mouse monoclonal antibody, DYS2 (Novocastra-Laboratories) raised against the last 17 amino acids of dystrophin (*a*, lanes 1, 2 and *b*), and a rabbit polyclonal antibody raised against dystrophin of *Torpedo marmorata* electrocyte²³ (*a*, lanes 4, 5). The specific immune complexes were detected with either a rabbit anti-mouse immunoglobulin antiserum, or a swine anti-rabbit immunoglobulin antiserum, both conjugated with horseradish peroxidase.

rare β -galactosidase-positive fibres observed in the contralateral, non-injected muscle when the recombinant Ad-RSV β gal was used for the injection (our unpublished data). This result validates the use of the contralateral non-injected muscle as a negative control in subsequent experiments.

Immunofluorescence studies were done on cryosections of the injected muscle, and results were compared to those obtained in a normal mouse muscle and in the contralateral non-injected muscle (Fig. 3). We observed a sarcolemmal immunostaining in 5 to 50% of fibres (Table 1) in the injected muscles, using a polyclonal antibody raised against the dystrophin from *Torpedo marmorata* electrocyte²³ (Fig. 3*b, d*). Such a variability in the results could be, at least partly, ascribed to the viral stock used and its titre (Table 1). In contrast, using the same antibody, less than 1% of fibres were labelled in the contralateral non-injected muscle (Fig. 3*c*). It has been established that few dystrophin-positive revertant fibres can be detected in *mdx* mouse muscles²⁴. We further verified with P20 AB²⁵, a polyclonal antibody raised against epitopes absent in the minidystrophin, that most of the positive fibres observed in the injected muscle were not revertants. Indeed, less than 1% of fibres were labelled with P20 AB

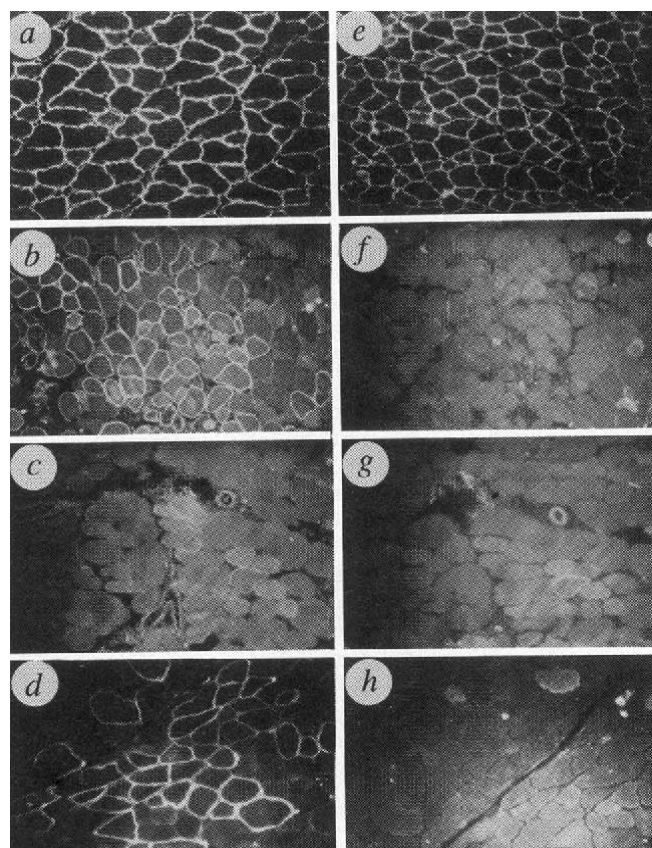


FIG. 3 Cryosections of normal mouse muscle (*a, e*), injected *mdx* muscles, 2 weeks (*b, f*) and 3 months after injection (*d, h*), contralateral non-injected muscle (*c, g*), immunostained with two polyclonal antibodies: a rabbit polyclonal antibody raised against dystrophin of *Torpedo marmorata* electrocyte²³, recognizing both minidystrophin and the full-length dystrophin (*a-d*) and P20 AB²⁵, a rabbit polyclonal antibody raised against the region covering amino-acid sequence 1,749-2,248 of the human dystrophin, recognizing only the full-length dystrophin (*e-h*).

METHODS. Unfixed cryosections ($10 \mu\text{m}$) of muscles were first incubated for 1 h at 37°C with the primary antibody diluted $1/50$ in PBS (pH 7.2), 0.3% BSA, washed $3 \times$ in PBS, 0.1% Tween 20, then incubated at room temperature with a fluorescein isothiocyanate-conjugated anti-rabbit immunoglobulin (SIGMA) for 1 h , washed $3 \times$ in PBS and mounted in 13% Mowiol 4-88 (Hoechst), 33% glycerol, 0.13 M Tris (pH 8.5). The photomicrographs were obtained with a Leitz epifluorescence microscope.

in the injected (Fig. 3f, h) as well as in the contralateral non-injected muscles (Fig. 3g).

In some experiments, the 3–4-mm-long injected biceps femoris muscle was sectioned in totality and the different sections were examined by immunofluorescence. In these cryosections, the same proportion of positive fibres was observed throughout the muscle, indicating that the Ad vector diffuses in the muscle from the injection point, at least to a certain extent. One animal (number 138) was killed 13 weeks after injection. In the injected muscle of this animal, we observed 20% of minidystrophin-positive fibres (Fig. 3d). This result is in the same range as those obtained in sibling mice injected with the same virus stock and killed 2 weeks (number 121) and 4 weeks (number 125) after injection, respectively (Table 1, Fig. 3b).

These experiments demonstrate that the recombinant Ad-RSVmDys can efficiently direct the *in vivo* synthesis of a significant amount of a minidystrophin, with a correct sarcolemmal localization, in the injected muscle. The expression of the transferred gene seems to be stable at least for 3 months. Whether or not this expression could correct the necrosis-regeneration process^{15,26} remains to be determined. These results must be compared to those obtained by direct injection of naked DNA. In this latter case, the expression was limited to about 2% of fibres⁸. In fact, the mechanism of uptake of naked DNA by fibres remains unclear and seems to be much less efficient than Ad vector-mediated gene transfer²⁷.

Adenoviruses have several interesting properties. They are: (1) stable and grow to very high titres in culture; (2) able to infect post-mitotic cells, such as myotubes; (3) able to infect many tissues. They could thus be used for the systemic delivery of foreign DNA. This would be of particular interest in DMD that affects all muscles, as the prognosis for this disease depends mainly on the functional impairment of the heart and respiratory muscles, which are not accessible to local injections. Although recombinant Ad vectors seem to remain in an extrachromosomal location and are replication-defective, recent reports have demonstrated that the expression of foreign genes transferred by this means could be observed for long periods especially in liver⁹ and skeletal muscles¹⁵. Among viral vector candidates for human gene therapy, adenoviruses are therefore particularly attractive.

We have demonstrated that it is possible to transfer a functional minidystrophin gene to a significant proportion of muscle fibres by intramuscular injection of an Ad vector. Further investigations will determine whether a similar result can be obtained using systemic administration and whether this results in functional improvement of dystrophin-deficient muscles. □

Received 5 November; accepted 14 December 1992.

- Emery, A. E. H. *Duchenne Muscular Dystrophy* (Oxford Univ. Press, Oxford, 1988).
- Zubrzycka-Gaarn, E. E. et al. *Nature* **333**, 466–469 (1988).
- Campbell, K. P. & Kahl, S. D. *Nature* **338**, 259–262 (1989).
- Koenig, M. et al. *Cell* **50**, 509–517 (1987).
- Partridge, T. A., Morgan, J. E., Coulton, G. R., Hoffman, E. P. & Kunkel, L. M. *Nature* **337**, 176–179 (1989).
- Huard, J. et al. *Muscle Nerve* **15**, 550–560 (1992).
- Gussoni, E. et al. *Nature* **356**, 435–438 (1992).
- Acsadi, G. et al. *Nature* **352**, 815–818 (1991).
- Stratford-Perricaudet, L. D. et al. *Hum. Gene Ther.* **1**, 241–256 (1990).
- Rosenfeld, M. A. et al. *Science* **252**, 431–434 (1991).
- Rosenfeld, M. A. et al. *Cell* **68**, 143–155 (1992).
- Quantin, B., Perricaudet, L. D., Tajbakhsh, S. & Mandel, J. L. *Proc. natn. Acad. Sci. U.S.A.* **89**, 2581–2584 (1992).
- Stratford-Perricaudet, L. D., Makeh, I., Perricaudet, M. & Briand, P. *J. clin. Invest.* **90**, 626–630 (1992).
- England, S. B. et al. *Nature* **343**, 180–182 (1990).
- Bulfield, G., Siller, W. G., Wight, P. A. L. & Moore, K. J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1189–1192 (1984).
- Sicinski, P. et al. *Science* **244**, 1578–1580 (1989).
- Hoffman, E. P. & Kunkel, L. M. *Neuron* **2**, 1019–1029 (1989).
- Sherk, T. & Williams, J. in *Current Topics in Microbiology and Immunology* (ed. Doerfler, W.) 1–39 (Springer, Berlin, 1984).
- Gluzman, Y., Reich, H. & Solnick, D. in *Eukaryotic Viral Vectors* (ed. Gluzman, Y.) 187–192 (Cold Spring Harbor Laboratory Press, New York, 1982).
- Ragot, T., Eloit, M. & Perricaudet, M. in *Human Gene Transfer* (eds Cohen-Haguenauer, O. & Boiron, M.) 249–260 (John Libbey, London, 1991).
- Graham, F. L., Smiley, J., Russell, W. C. & Nairn, R. *J. gen. Virol.* **36**, 59–72 (1977).
- Levero, M. et al. *Gene* **101**, 195–201 (1991).

- Cartaud, A. et al. *Neuroscience* **48**, 995–1003 (1992).
- Hoffman, E. P., Morgan, J., Watkins, S. C. & Partridge, T. A. *J. neurol. Sci.* **99**, 9–25 (1990).
- Ginjaar, H. B. et al. *FEBS Lett.* **308**, 293–297 (1992).
- Stedman, H. et al. *Nature* **352**, 536–539 (1991).
- Jiao, S. et al. *Hum. Gene Ther.* **3**, 21–33 (1992).
- Hugnot, J. P. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 7506–7510 (1992).

ACKNOWLEDGEMENTS. We thank K. Davies (Molecular Genetics Group, John Radcliffe Hospital, Oxford, UK) for providing the Becker-like cDNA, G. J. van Ommen and H. B. Ginjaar (Sylvius Laboratory, Leiden, The Netherlands) for P20 antibody, C. Chianale (Laboratoire d'Expérimentation Animale of the Institut Gustave Roussy) for *mdx* mice breeding and nursing, and A. Grégoire for her help in preparation of the virus stocks. We acknowledge the support of the Association Française contre les Myopathies, the CNRS, the Université Paris V and the Ministère de la Recherche et de l'Espace. T.R. and N.V. are fellows of the Association Française contre les Myopathies.

Use of evolutionary limitations of HIV-1 multidrug resistance to optimize therapy

Yung-Kang Chow, Martin S. Hirsch, Debra P. Merrill, Lawrence J. Bechtel, Joseph J. Eron, Joan C. Kaplan & Richard T. D'Aquila

Infectious Disease Unit, Gray 5, Massachusetts General Hospital & Harvard Medical School, Boston, Massachusetts 02114, USA

WILD-TYPE reverse transcriptase has evolved for the survival of human immunodeficiency virus type 1 (HIV-1) by natural selection¹. In contrast, therapy relying on inhibitors of reverse transcriptase by nucleosides like zidovudine (AZT) or dideoxyinosine (ddI), and by non-nucleosides like pyridinones or nevirapine^{2–6}, may exert different selection pressures on this enzyme. Therefore the acquisition of resistance to reverse transcriptase inhibitors by selection of mutations in the *pol* gene^{7–15} may require compromises in enzyme function that affect viral replication. As single mutations are unlikely to confer broad resistance when combinations of reverse transcriptase inhibitors are used, multiple mutations may occur that result in further compromises. Certain drug combinations may prevent the co-existence of adequate reverse transcription function and multi-drug resistance (MDR). Unlike bacterial or eukaryotic drug resistance, retroviral drug resistance is conferred only by mutations in its own genome¹⁶ and is limited by genome size. Combining drugs directed against the same essential viral protein may thus prevent HIV-1 MDR, whereas the conventional approach of targeting different HIV-1 proteins for combination therapy may not, because genomes with resistance mutations in different HIV-1 genes might recombine to develop MDR¹⁷. Here we show that several mutations in the HIV-1 reverse transcriptase gene that confer resistance to inhibitors of this enzyme can attenuate viral replication. We tested whether combinations of mutations giving rise to single-agent resistance might further compromise or even abolish viral replication, and if multidrug-resistant viruses could be constructed. Certain combinations of mutations conferring resistance to AZT, ddI and pyridinone are incompatible with viral replication. These results indicate that evolutionary limitations exist to restrict development of MDR. Furthermore, a therapeutic strategy exploiting these limitations by using selected multidrug regimens directed against the same target may prevent development of MDR. This approach, which we call convergent combination therapy, eliminated HIV-1 replication and virus breakthrough *in vitro*, and may be applicable to other viral targets. Moreover, elimination of reverse transcription by convergent combination therapy may also limit MDR.

To prove that resistance to reverse transcriptase inhibitors may compromise viral replication, provirus mutants containing multiple drug resistance mutations in the enzyme were constructed. Mutant provirus DNAs were transfected into COS-7 cells, and transfection supernate virus stocks were used to infect T cells. Mutations conferring resistance to ddI, AZT and non-nucleoside inhibitors were studied (see Fig. 1 legend)^{7–9}.

Quantitative virus yield assays indicated that more infectious virions were produced after infections of PHA-stimulated peripheral blood mononucleocytes (PBMC) with wild-type virus than with mutants 2 or 3 (Fig. 1). Reverse transcriptase activity and p24 levels in the supernates were also higher in PBMC cultures infected with wild-type HIV-1 than with mutants 2 or 3 (Figs 1 and 2). Similar results were obtained during infections of T cell lines. Quantitation by polymerase chain reaction (PCR) of HIV-1 DNA after infection of T cells confirmed that infections with mutants 2 and 3 produced less reverse-transcribed HIV-1 DNA than did wild-type virus (Fig. 3a)¹⁸. Moreover, mutants 2 and 3 produced lower virion-associated reverse transcriptase activity than wild-type virus (Fig. 3b). We conclude that defects in the mutant virion-associated enzyme activity were responsible for the attenuated replication, although the precise defects remain to be determined. The observation that several HIV-1 AZT-resistance mutations do not persist following withdrawal of drug-selection pressure *in vivo* is consistent with this reduced replication phenotype of mutant viruses^{8,19-21}.

HIV-1 may develop MDR by combining mutations causing resistance to single agents. For example, one combination mutant gave 6-fold and 18-fold decreases in sensitivity to AZT and ddI respectively⁸. But, with the accumulation of several mutations, some combinations might compromise enzymatic functions so severely as to render HIV-1 replication incompetent. Such incompatibility among drug-resistance mutations would suggest that evolutionary limits restrict MDR development. The observation that resistance mutations cluster around selected regions in the enzyme is consistent with such constraints^{7-15,22}.

When an additional non-nucleoside resistance mutation was added to mutant 3, reverse transcriptase activity became undetectable, although p24 levels in transfectant supernates were unaffected (Fig. 3b). This quadruple mutant (mutant 4) was non-infectious, as indicated by the lack of cytopathic effects,

supernate reverse transcriptase activity, and p24 in T cell cultures (Fig. 2). Furthermore, only minimal levels of HIV-1 DNA were detected 24 to 72 hours after infection (Fig. 3a), and no HIV-1 DNA was detectable by PCR 21 days after infection. These results indicate that mutant 4 does not establish productive infections, and that simultaneous presence of these resistance mutations at codons 74, 103, 215 and 219 is incompatible with virus replication.

PCR experiments with this replication-incompatible mutant suggest that reverse transcriptase activity was greatly diminished during infection (Fig. 3a). Moreover, the absence of detectable transfection supernate enzyme activity indicates that activity is minimal in mutant virions (Fig. 3b). Thus the cumulative effects of these mutations may have suppressed the polymerase activity below the threshold necessary for viral replication²². These observations suggest that certain convergent combination regimens using reverse transcriptase inhibitors may prevent MDR if the structural demands for broad resistance are incompatible with threshold activity of the enzyme.

Having shown that this combination of resistance mutations for AZT, ddI and for non-nucleosides is incompatible, we tested whether the convergent combination of AZT, ddI, and pyridinone was effective in both prophylaxis and treatment of HIV-1 infections. Viral breakthrough was assessed by cytopathic effects, p24 production, and quantitative PCR. After 5 hours of drug pretreatment of cells, we found that the combination of AZT and HIV protease inhibitor plus recombinant interferon- α A did not prevent HIV-1 infection (Fig. 4a). Two-drug combinations of AZT plus ddI or AZT plus pyridinone were similarly ineffective, but AZT plus ddI and pyridinone prevented viral replication. When drugs were removed 21 days after infection, no viral replication was observed for the next 30 days. Furthermore, HIV-1 DNA was not detected in this culture using PCR conditions that detect a single copy of HIV-1 DNA (Fig. 4a).

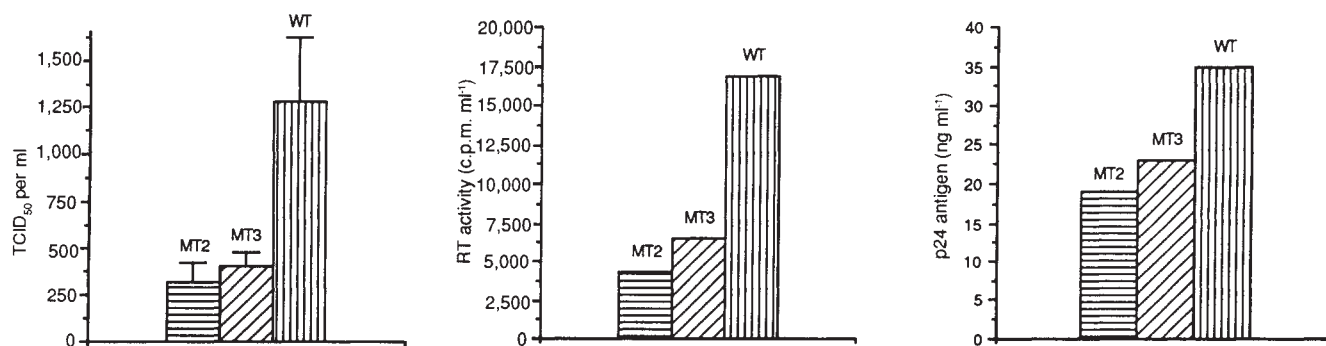


FIG. 1 Infectious virus yield, supernatant fluid reverse transcriptase (RT) activity and HIV-1 p24 antigen levels in phytohaemagglutinin (PHA)-stimulated PBMC cultures infected with transfectant viruses.

METHODS. Site-directed mutagenesis was done on a genetic background of pSP65-HXB2-gpt^{25,26}, followed by sequence confirmation of the entire RT gene for all mutant molecular provirus clones²⁷. Drug-resistance RT mutations analysed here include: Thr 215 → Tyr and Lys 219 → Gln (AZT resistance)⁷, Leu 74 → Val (ddI resistance)⁸, and Lys 103 → Asn (conferring resistance to non-nucleoside inhibitors like pyridinone⁹ and nevirapine). The following HXB2-derived RT mutant viruses were studied: MT2 (Thr 215 → Tyr and Lys 219 → Gln) and MT3 (Thr 215 → Tyr, Lys 219 → Gln and Leu 74 → Val). Transfection of HIV-1 provirus DNA into COS-7 cells was done using a DEAE-dextran method²⁸. Transfectant virus present in the culture supernatant fluid was titrated using end-point dilution by scoring of cytopathic effects (CPE) on C8166 cells²⁹. PBMC infections with these transfectant virus stocks used 250 tissue-culture infectious doses (TCID₅₀) per million cells. Wild-type virus was compared with mutants 2 and 3 with respect to infectivity, RT activity and p24 antigen levels of output viruses produced from such infections. Infectious output viruses were titrated by CPE in sextuplet on C8166 cells 10 days after infection (left panel). Geometric

means and standard errors calculated using the Spearman-Kärber method are shown. The 95% confidence interval for wild-type (WT): 804–2035 TCID₅₀ per ml; mutant 2 (MT2): 185–552 TCID₅₀ per ml; and mutant 3 (MT3): 280–580 TCID₅₀ per ml (WT versus MT2, $P=0.002$; WT versus MT3, $P=0.02$, Student's *t*-test). Supernatant fluid PEG-precipitable RT activity (middle panel) and p24 antigen (right panel) were determined at the same time. RT assay and p24 antigen ELISA were done as described²⁹. Day-7 results (not shown) were similar to day-10 results. We also studied mutant 74 (Leu 74 → Val), mutant 3–103 (Thr 215 → Tyr, Lys 219 → Gln and Lys 103 → Asn), and mutant 103 (Lys 103 → Asn), which conferred 20-fold increases in IC₅₀ over WT to both nevirapine and pyridinone (data not shown). All of the mutants described above were replication-competent in T cells with detectable RT activity and p24 antigen in transfectant culture supernatant fluids. The mean 50% inhibitory concentration (IC₅₀) values for WT are AZT, 0.002 μ M; ddI, 0.8 μ M; for MT2, they are: AZT, 0.017 μ M; ddI, 1.1 μ M; and for MT3 are: AZT, 0.004 μ M, ddI, 4.0 μ M (J.J.E. *et al.*, manuscript in preparation). For Figs 2 and 3, mutant 4 (MT4) contained all the substitutions in mutant 3 (Thr 215 → Tyr, Lys 219 → Gln and Leu 74 → Val) plus Lys 103 → Asn. Other mutants with additional drug resistance substitutions have also been shown to be replication-incompetent (data not shown).

The same convergent combination was then tested in ongoing infections (Fig. 4b). Drugs were added to PBMC cultures at peak HIV-1 production (310 ng HIV-1 p24 antigen per ml supernate) 7 days after infection. The combination of AZT with interferon plus recombinant soluble CD4 (rsCD4)²³ did not suppress HIV-1 breakthrough (Fig. 4b). In contrast, 35 days after convergent combination treatment, p24 was undetectable (Fig. 4b). Furthermore, HIV-1 DNA was not detectable 21 days after beginning drug treatment. Even after drugs were removed at day 56, no viral replication was observed for the next 45 days (Fig. 4b).

The drug concentrations used here in the successful treatment of established infection (0.3 μ M AZT, 10 μ M ddI, and 0.09 μ M pyridinone) are easily attainable in plasma (achievable plasma concentrations are 4 μ M AZT, 10 μ M ddI²⁴, and 1 μ M pyridinone L697,661; E. Emini, personal communication). It is unclear whether the virus breakthrough suppression results from limitations in resistance development. But as HIV-1 replication can also be abrogated using other convergent combinations of inhibitors, such as AZT plus ddI and foscarnet, or AZT plus ddI and nevirapine (Y.-K.C. *et al.*, manuscript in preparation), breakthrough suppression is not unique to a particular convergent combination. Regardless of the mechanism of breakthrough suppression, as reverse transcription is necessary to create HIV-1 variants, complete inhibition of this process using convergent

combination regimens should prevent emergence of multiply resistant viruses as well as any immune escape mutant viruses.

Furthermore, after ten PBMC passages of a clinical HIV-1 isolate with AZT- and ddI-selected mutations in AZT+ddI+nevirapine, we were unable to select for viruses triply resistant to AZT+ddI+nevirapine. We also failed to select for nevirapine resistance after ten PBMC passages in nevirapine using this isolate. In contrast, using an AZT- and ddI-sensitive isolate, highly nevirapine-resistant viruses emerged after 10 PBMC passages in nevirapine (Y.-K.C. *et al.*, manuscript in preparation). The difficulty in selecting for nevirapine resistance in a

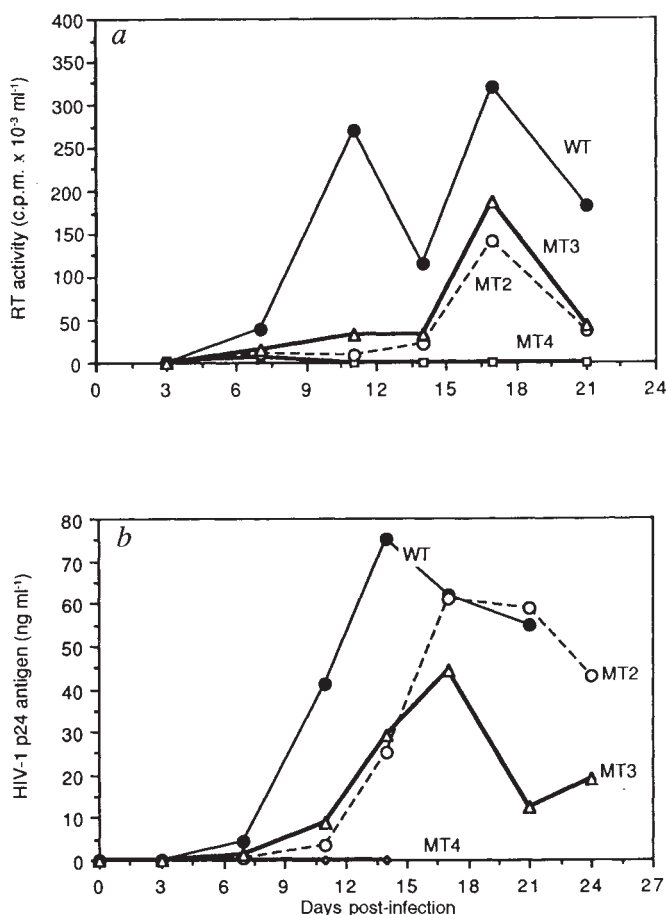


FIG. 2 Virus replication kinetics in PHA-stimulated PBMC. Wild-type and mutant viruses derived from transfection supernatant fluids were used to infect PBMC. Transfection-derived viruses (equivalent to 0.4 ng of p24 antigen) were used to infect 10×10^6 cells. HIV-1 replication was monitored by levels in the supernatant fluid of a, RT activity, and b, p24 antigen levels. RT assay and p24 antigen ELISA were done as described²⁹. Similar results were obtained in experiments using different quantities of virus (0.5 ng p24 antigen and 5×10^6 cells) or in experiments using H9, MT-2 and SupT1 cells.

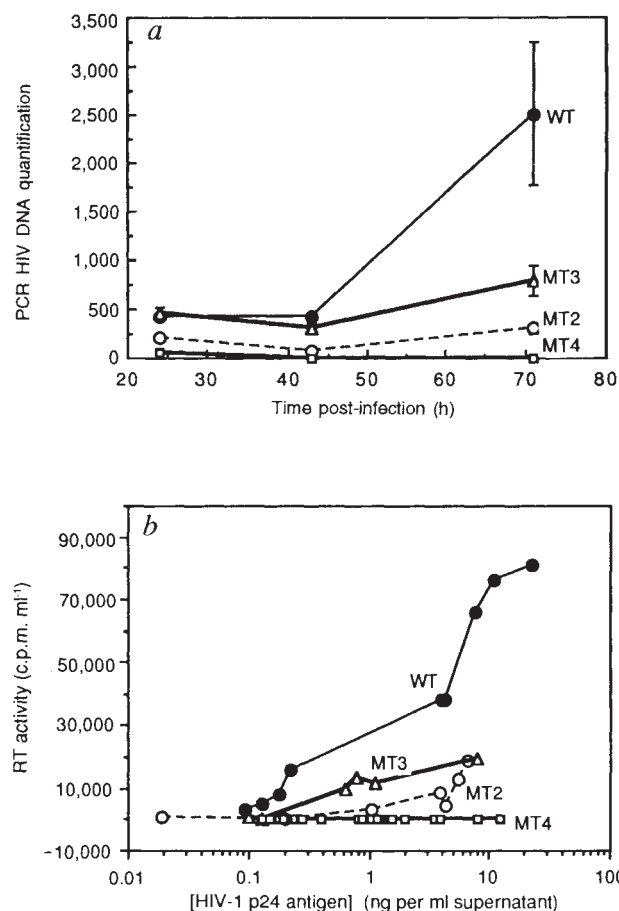
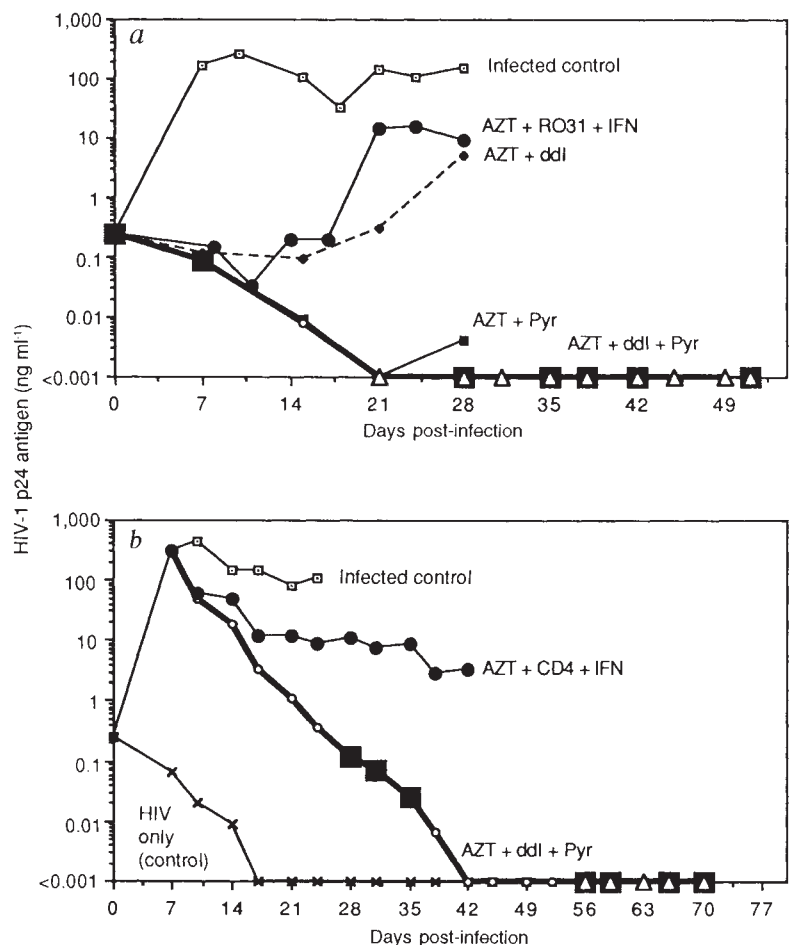


FIG. 3 Quantitation of RT activity during infection and virion-associated RT activity of viruses derived from COS-7 cell transfections. a, PCR DNA quantification of reverse-transcribed HIV-1 DNA in infected SupT1 cells. Transfectant virus stocks were first treated with 30 U DNase (Promega) and 1 mM MgCl_2 for 40 min at 25 $^{\circ}\text{C}$, conditions previously shown not to affect infectivity²³. Infection was with 0.5 ng p24 antigen-equivalent of virus per 10^6 cells. At various times after infection, cells were collected, counted and lysed at 1.2×10^4 cells per μl . PCR HIV-1 DNA was quantified as described¹⁸. Copy numbers of HIV-1 genomes in 1.2×10^5 cell equivalents were determined by reference to a standard curve and were plotted against time after infection. Values are averages of 13–14 determinations in SupT1 cells or 6–7 determinations in H9 cells (not shown). The standard error of the mean is shown for each time point. Infections were monitored from 24–72 h. No differences were observed in other experiments when infections were monitored 6 to 120 h post-infection. Similar results were obtained for mutants 2 and 3 on day 21 post-infection in experiments using 0.04–0.1 ng p24 antigen-equivalent viruses on SupT1 cells and PBMC. In those experiments, HIV-1 DNA was not detectable for mutant 4 (not shown). b, Quantitation of virion-associated RT activity of viruses derived from transfections. The levels of PEG-precipitable RT activity present in virions derived from COS-7 transfectant viruses were plotted against virus particle numbers approximated by p24 antigen levels. Transfection, RT assay and p24 antigen ELISA were done as described^{28,29}.

FIG. 4 Prophylaxis against HIV-1 infection and elimination of breakthrough virus replication in established HIV-1 infections of PBMC cultures by convergent combinations of RT inhibitors. *a*, Prevention of HIV-1 replication and breakthrough. *b*, Elimination of virus replication and eradication of HIV-1 from established infections of PBMC cultures.

METHODS. *a*, PHA-stimulated PBMC were pretreated for 5 h with 95% inhibitory concentrations (IC_{95}) of the drugs: AZT, 0.1 μ M; ddI, 4 μ M; pyridinone L697,661 (Pyr), 0.12 μ M; HIV-1 protease inhibitor RO31-8959 (RO31), 0.024 μ M. A similarly effective concentration of recombinant interferon- α (IFN, 32 U ml^{-1}) was used. Cells were infected with HIV-1 in the presence of the drugs, passaged biweekly and maintained with drug-containing medium through 28 days in culture²³. After 21 days in culture, one-third of the cells and the supernatant fluid in the (AZT + ddI + pyridinone)-treated culture was removed from the drugs, resuspended in drug-free medium (triangles), and followed for 30 d for CPE, p24 antigen production (y axis), and PCR HIV-1 DNA quantitation. HIV-1 DNA PCR-negative sample (40 cycles of amplification) are designated by solid squares. The AZT + ddI + pyridinone culture was negative for HIV-1 DNA by day 7. Similar results were seen when p24 values were adjusted for viable cell numbers. *b*, PHA-stimulated PBMC were infected with HIV-1 at a multiplicity of infection of 1,000 TCID per 10^6 cells (0.246 ng ml^{-1} p24 antigen equivalent of virus). Drugs were added at the day of peak infection (day 7 after infection) to the infected cultures (310 ng ml^{-1} p24 antigen equivalent of HIV-1 in culture supernatant fluid). AZT and ddI were used at 99% inhibitory concentration (AZT, 0.3 μ M; ddI, 10 μ M); pyridinone L697,661 (Pyr) was used at 50% inhibitory concentration (0.09 μ M); rs CD4 (CD4) and recombinant interferon- α (IFN) were used at concentrations of 0.32 μ g ml^{-1} and 32 U ml^{-1} respectively. Cell cultures were subsequently passaged biweekly and maintained with drug-containing medium through 73 days in culture²³. After 56 days in culture (49 days after drug treatments), one-third of the cells and the supernatant fluid in the (AZT + ddI + pyridinone)-treated culture was resuspended in drug-free medium (triangles), and followed for 45 d for CPE, p24 antigen production (y axis), and PCR HIV-1 DNA quantitation. HIV-1 DNA PCR-negative samples (40 cycles of amplification) are designated by solid squares. HIV-1 DNA and p24 were no longer detectable by days 28 and 42 after infection respectively (days 21 and 35 after drug treatments), and the culture remained negative for HIV-1 for the remainder of the 101 day duration of the experiment. In *a* and *b*, input DNA was removed by incubating virus stocks with DNase (Promega) at 10 U ml^{-1} and $MgCl_2$ at 1 mM for 45 min at 25 °C, conditions that do not affect infectivity¹⁸. PHA-stimulated PBMC were infected with HIV-1 14a Pre, a drug-sensitive clinical isolate of HIV-1 from an untreated patient³⁰, at a multiplicity of infection of 1,000 TCID per 10^6 cells (0.246 ng ml^{-1} p24 antigen equivalent of virus). Virus was titrated



using end-point dilution assays by CPE scoring on PBMC. The IC_{50} of 14a Pre are: AZT, 0.00028–0.01 μ M; ddI, 0.67 μ M³⁰; pyridinone L697,661, 0.09 μ M. Fresh uninfected donor PBMC were added every week to replenish aged cultures. Culture flasks were monitored for CPE, and supernates collected for p24 antigen determinations and PCR quantitation of HIV-1 DNA in 10^5 infected cells¹⁸. Forty cycles of amplification were used for PCR¹⁸, a condition that detects a single copy of control HIV-1 DNA. Drug toxicity controls were run in parallel on uninfected PBMC and H9 cells; no cell toxicity was observed at the concentrations used.

genetic background of AZT- and ddI-selected mutations strongly supports the hypothesis that MDR may be evolutionarily limited by convergent combination regimens.

To our knowledge, these are the first demonstrations of a replication-attenuated phenotype for *in vivo* drug-selected HIV-1 mutants and of replication-incompatible combinations of drug resistance mutations in any viral system. But *in vivo* attenuation may not be synonymous with decreased virulence. After many rounds of replication, even small replication defects may translate into major differences in virus production rate. Other mutations could also impair HIV-1 replication, and even if certain combinations of drug-resistance mutations in reverse transcriptase do not completely attenuate replication, they may not necessarily confer MDR. For example, mutant 3 is resistant to ddI but sensitive to AZT (Fig. 1 legend). This potential loss of resistance phenotypes of individual mutations in a combination mutant, as might happen after a recombination event, poses additional limitations to MDR exploited by convergent combination therapy.

The approach used here can be applied to optimize combinations of antiretroviral agents, although extrapolation from *in vitro* experiments to clinical therapy may be difficult, and MDR may develop against some convergent combinations. Neverthe-

less, our results suggest that convergent combination therapy may be beneficial in the treatment of HIV-1 infections and in post-exposure prophylaxis and this will be investigated in clinical trials. □

Received 30 July; accepted 7 December 1992.

1. Temin, H. M. *J. AIDS* **2**, 1–9 (1989).
2. Pauwels, R. *et al. Nature* **343**, 470–474 (1990).
3. Meriuzzi, V. J. *et al. Science* **250**, 1411–1413 (1990).
4. Goldman, M. E. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 6863–6867 (1991).
5. Romero, D. L. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 8806–8810 (1991).
6. Balzarini, J. *et al. Antimicrob. Ag. Chemother.* **36**, 1073–1080 (1992).
7. Larder, B. A. & Kemp, S. D. *Science* **246**, 1155–1158 (1989).
8. St. Clair, M. H. *et al. Science* **253**, 1557–1559 (1991).
9. Nunberg, J. H. *et al. J. Virol.* **65**, 4887–4892 (1991).
10. Richman, D. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 11241–11245 (1991).
11. Prasad, V. R., Lowy, I., De Los Santos, T., Chiang, L. & Goff, S. P. *Proc. natn. Acad. Sci. U.S.A.* **88**, 11363–11367 (1991).
12. Mellors, J. W. *et al. Molec. Pharmacol.* **41**, 446–451 (1992).
13. Kellam, P., Boucher, C. A. B. & Larder, B. A. *Proc. natn. Acad. Sci. U.S.A.* **89**, 1934–1938 (1992).
14. Gao, Q., Gu, Z., Parriak, M. A., Li, X. & Wainberg, M. A. *J. Virol.* **66**, 12–19 (1992).
15. Fitzgibbon, J. E. *et al. Antimicrob. Ag. Chemother.* **36**, 153–157 (1992).
16. Coen, D. M. *J. Antimicrob. Chemother.* **18**, Suppl. B 1–10 (1986).
17. Hu, W.-S. & Temin, H. M. *Science* **250**, 1227–1233 (1990).
18. Eron, J. J., Gorczyca, P., Kaplan, J. C. & D'Aquila, R. T. *Proc. natn. Acad. Sci. U.S.A.* **89**, 3241–3255 (1992).
19. Boucher, C. A. B. *et al. J. infect. Dis.* **165**, 105–110 (1992).
20. Albert, J. *et al. J. Virol.* **66**, 5627–5630 (1992).
21. Land, S., McGavin, C., Lucas, R. & Birch, C. *J. infect. Dis.* **166**, 1139–1142 (1992).

22. Larder, B. A., Kemp, S. D. & Purifoy, D. J. M. *Proc. natn. Acad. Sci. U.S.A.* **86**, 4803-4807 (1989).
23. Johnson, V. A., Barlow, M. A., Merrill, D. P., Chou, T.-C. & Hirsch, M. S. *J. infect. Dis.* **161**, 1057-1067 (1990).
24. Yarchoan, R., Mitsuya, H., Myers, C. E. & Broder, S. *New Engl. J. Med.* **321**, 726-738 (1989).
25. Kunkel, T. A., Roberts, J. D. & Zakour, R. A. *Meth. Enzym.* **154**, 367-382 (1987).
26. Ratner, L. *et al. AIDS Res. hum. Retrovir.* **3**, 57-69 (1987).
27. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
28. Kornfeld, H., Riedel, N., Viglianti, G. A., Hirsch, V. & Mullins, J. I. *Nature* **326**, 610-613 (1987).
29. Aldovini, A. & Walker, B. *Techniques in HIV Research* (Stockton, New York, 1990).
30. Johnson, V. A. *et al. J. infect. Dis.* **164**, 646-655 (1991).

ACKNOWLEDGEMENTS. We thank L. Ratner for pSP65-HXB2-gpt, E. Emini and M. Goldman of Merck for pyridinone L697 661, V. J. Merluzzi of Boehringer Ingelheim for nevirapine, I. Sim of Hoffmann-La Roche for rIFN- α A and N. Roberts of Roche Products, UK for R031-8959 protease inhibitor, and Biogen for rSCD4; B. Walker, S. Rusconi, T. Mazzulli, D. Manion, D. Zhang, A. Caliendo, D. Weissman, T. and E. Harter for critically reviewing the manuscript; and D. Strick, B. Smith, K. DeVore and R. Byington for assistance. This work was supported by NCI, NIAID and MSTP.

Transcriptional differences in polymorphic and conserved domains of a complete cloned *P. falciparum* chromosome

Michael Lanzer, Derik de Bruin & Jeffrey V. Ravetch

DeWitt Wallace Research Laboratory, Sloan-Kettering Institute, Division of Molecular Biology, 1275 York Avenue, New York, New York 10021, USA

CLASSICAL genetic studies on the human malaria parasite *Plasmodium falciparum* have been hampered by a complex life cycle which alternates between vertebrate and invertebrate hosts. Consequently, only a few genetic crosses have been performed so far¹⁻⁴. In addition, molecular genetics has provided only limited access to the genes of this pathogen, a consequence of an unusually high A + T content^{5,6}. To overcome these limitations we have constructed an ordered telomere-to-telomere contig map of *P. falciparum* chromosome 2 by isolating overlapping yeast artificial chromosome clones. This approach was used to examine the strain-dependent polymorphisms commonly observed for *P. falciparum* chromosomes^{7,8}. Our analysis reveals that polymorphisms of chromosome 2 are restricted to regions at either end, representing 20% of the chromosome. Transcription mapping of the entire chromosome suggests a compartmentalization of chromosome 2 into a transcribed central domain and silent polymorphic ends.

A yeast artificial chromosome (YAC) library of the *P. falciparum* strain FCR3 (ref. 9) was screened by polymerase chain reaction (PCR) with probes to known chromosome 2 markers¹⁰. Remaining gaps were filled by rescuing the ends of the YAC clones, thereby defining new probes, sequence tag sites (STS), for screening the library. The *P. falciparum* telomere-containing YAC clones GC6 and CB4 were isolated from a YAC library of FCR3 enriched for subtelomeric DNA fragments⁹. Eighteen YAC clones and 31 STS markers were obtained (Fig. 1a, and Table 1). All STS, as well as the left end of both telomere YAC clones, hybridize back to chromosome 2 (data not shown). The integrity of all YAC clones was confirmed by restriction analysis in comparison to genomic FCR3 DNA (summarized in Fig. 1a). The YAC clones GC6 and CB4 hybridize to the *P. falciparum* rep20 sequence⁵, an element exclusively found in subtelomeric regions¹¹, and to an oligonucleotide specific for the *P. falciparum* telomere repeat sequence (Fig. 1b, top panels). For both telomeric YAC clones, the hybridization to the *P. falciparum* telomere oligonucleotide is sensitive to digestion with the exonuclease Bal-31 (Fig. 1b, bottom panel), indicating that the *P. falciparum* telomere sequences are terminal and had served as a substrate for the addition of yeast telomere sequences. Neither the rep20 sequence nor the telomere repeat oligonucleotide hybridized to any other YAC clone shown in Fig. 1a, with the exception of clone CG7, which overlaps substantially with CB4 and hybridizes to rep20 sequences (data not shown). These data

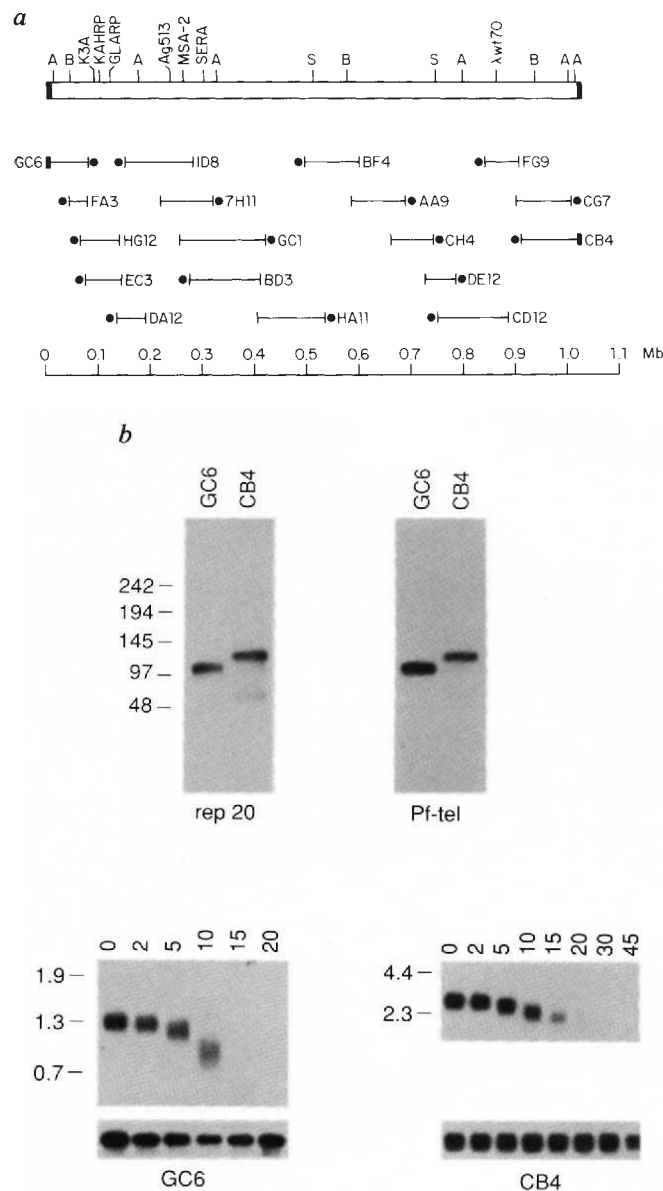


FIG. 1 a, Representative YAC contig of *P. falciparum* chromosome 2. The location of chromosome 2-specific genes and one anonymous marker are indicated. KAHRP, knob-associated histidine-rich protein²⁰; Ag513 (ref. 21); MSA-2, merozoite surface antigen 2 (ref. 22); SERA, serine-repeat antigen²³; λ wt70 (ref. 10). The genes K3A and GLARP (Gln-Lys-Asn-rich protein) will be described elsewhere (D. de B., M.L. and J.V.R., manuscript in preparation). Filled rectangles indicate the position of *P. falciparum* telomere repeat sequences. The cleavage sites of the restriction endonucleases *Apal* (A), *Bgl* (B) and *Sma* (S) are indicated on the chromosome 2 map. The left arm of each YAC clone is indicated by a filled circle. The two telomere YAC clones GC6 and CB4 contain only the left vector arm of pYAC4 (ref. 9). b, Characterization of the *P. falciparum* telomere YAC clones. Southern blots containing the two telomere YAC clones, GC6 and CB4, were hybridized with a probe to the rep20 sequence⁵ (top, left panel) and with a *P. falciparum* telomere repeat oligonucleotide, Pf-tel (GGGTTTA)₄ (top, right panel). Bottom, Digestion with the exonuclease Bal-31. Genomic DNA (30 μ g), prepared as described⁹ from the YAC clones GC6 and CB4 respectively, were incubated with 20 U exonuclease Bal-31 (New England Biolabs) in 250 μ l at 37 $^{\circ}$ C. Aliquots were withdrawn at the times indicated (min) and digested with restriction endonuclease *Rsa*I. Southern filters were hybridized with the *P. falciparum* telomere repeat oligonucleotide. As a control, Southern filters were hybridized with internal probes, STS 3 for GC6 and STS 31 for CB4. DNA size standards are indicated on the left in kb. METHODS. YAC clone DNA embedded in agarose plugs was size-fractionated by pulse-field gel electrophoresis using a Bio-Rad CHEF-DR11 system (pulse-field conditions: ramped pulse from 5 to 25 s over 20 h at 180 V, 1% LE agarose (FMC), 0.5 $\times$ TBE, at 14 $^{\circ}$ C).

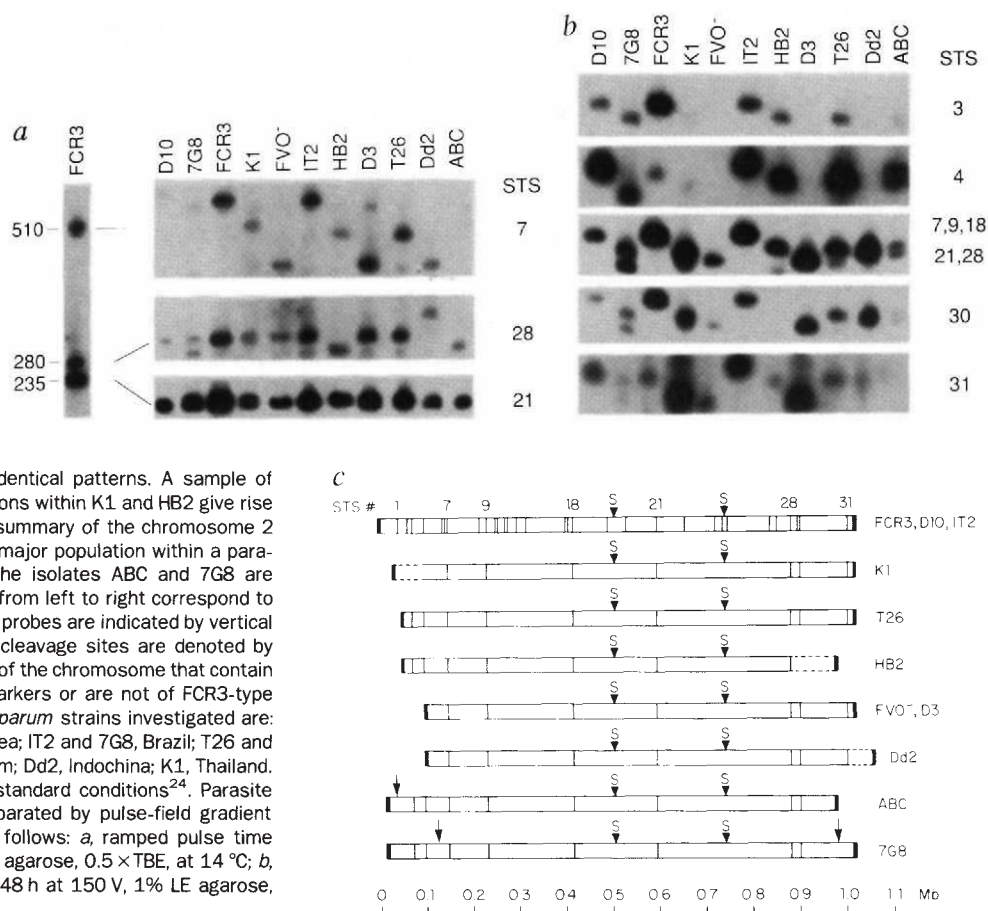
TABLE 1 Chromosome 2-specific sequence tag sites

STS	Origin	Primer 1	Primer 2	Length (bp)
1	FA3-LA	GTATTAGTAGTATTTCTAATAAA	TCTCCATTAAATTATACTATATA	520
2	HG12-LA	TTTTTGATTTATCAAAACCTC	GTATATAGAAATATATATTATAC	174
3	EC3-LA	CCTTTGTAAGTAACATAATCGGG	CGAAAGAACCTTCCCTTGGCT	750
4	KAHRP	GGAAACGGATCCGGTGACTCC	TACCATCGACAACATTTTCTCT	391
5	HG12-RA	GGAGGAATGTCTTTTATGTTC	GAAGGAAGAGTATGAAAATAAG	380
6	EC3-RA	ATTGCAAAAAATAGAATAATATG	GTTTTACACATAAACATTATGT	86
7	ID8-LA	ATCTGGTTTATCCAATCTAC	CATTTTGAAGATGATGACTAAA	250
8	7H11-RA	ACTTATCTATACAAACATGTG	TTACCTATGTACATGCTCTT	990
9	Ag513	CAGACGGTAAAGGAGAAGAG	CGTATACATAATCATGTGGTGAC	506
10	GC1-RA	CTAAGAATATGTAAATTATTG	TAAACAAGCTTTTCAGCTG	180
11	MSA-2	GCAACACATTACATAACAATGC	CATTTGATTAGTTTGAGAGTC	452
12	BD3-LA	AAAATCTGCTTACCAAAACC	CTAATATACTATTACACGAG	210
13	ID8-RA	GTGATAAAAAGAAATATACATATA	GTTTTTTTAAAGACTCAITTAAG	380
14	SERA	CAGGAGGAGGTCAAGCAGGTAATAC	TGAACCTGAAGTGAAGCTGAACCT	517
15	7H11-LA	CACAAAATAAAAGTTACACTTC	ACATCTTTTATTTTATTTAAGAC	220
16	HA11-RA	TACATAACAGTTGATATTATATA	ATCGTTACTTGTTTATATCTG	1,050
17	BD3-RA	CATATTTGGTTGTAAAGCCA	AAAAACGACGCATCTGTTTT	243
18	GC1-LA	GCCATTATGAAAAATAAATATAC	TTCTGTTCTTTTCTTTTCTTTC	148
19	BF4-LA	GGAATAATAACAAATGTGAATA	AGTTCCATATTGAATATATTCT	980
20	HA11-LA	CATAAGAATAGAACTTTAGTTG	CTTCTGTGTTCTTATTATTAC	370
21	BF4-RA	AACACACGTACACGCATAT	TTAAAAGTGTCATACCTCC	150
22	CH4-RA	TCAAGGCTATGGTTTAAACAG	TCAAGTCTTCTCGTACATT	1,200
23	DE12-RA	ATATTGTTGCGGTGGTTAA	GATATGGTTTATAAGATAAAAAG	520
24	CH4-LA	CCCATATCACTAAATAAAATAA	GAAGAGAGTTAATATAAATAAAG	180
25	CD12-LA	GTTCGATTGCTTTGTGTAT	CTAGTAATTATAAAGCTGAAG	300
26	FG9-LA	ND	ND	
27	λ wt70	CGGTTCTTCAGGAAACGTG	GGGTCTCATTTTAATAGGTG	580
28	CD12-RA	CGCATGAATTAATTTTGGTAA	TAATTATAAGGAACCGACATT	650
29	CG7-RA	TTTAGAATTTTTTATTTATAAAAAG	TTTTGTTTATATAACAGAAAAAG	300
30	FG9-RA	AAGTTTTTCATCTCCATTGG	ATGTTAAATCTTGTGAATTCG	520
31	CG7-LA	TTCTGGGTCACCTACTATAT	CCCCCTATGAAAAATTC	103

The origins of the chromosome 2-specific STS probes are indicated, with LA referring to the left end and RA to the right end of the corresponding YAC clone. The chromosomal locations of these STS probes are summarized in Fig. 2c. Primers are oriented from 5' to 3'. STS markers 1 and 31 are repetitive and hybridize to both ends of chromosome 2 as well as to other chromosomes (data not shown). The conditions for PCR amplification are: 35 cycles for 1 min at 94 °C, 2 min at 50 °C, 3 min at 72 °C, using a Perkin Elmer Cetus DNA Thermal Cycler. ND, Not determined.

FIG. 2 Analysis of chromosome 2 polymorphisms. *a*, Genomic DNA from various parasite strains was digested with the restriction endonuclease *Sma*I, size-fractionated by pulse-field gel electrophoresis and transferred to nitrocellulose. The Southern filter was sequentially probed with STS 7, 21 and 28. A composite of these data is shown. Subpopulations are observed in all the strains investigated. As a standard, a Southern filter containing *Sma*I-digested FCR3 DNA was hybridized simultaneously with STS probes 7, 21 and 28. DNA size markers are indicated on the left in kb. *b*, Separated whole chromosomes from various parasite isolates were hybridized with the STS probes indicated. Probes 7, 9, 18, 21, 28 were individually hybridized and gave identical patterns. A sample of these data is shown. FCR3-type subpopulations within K1 and HB2 give rise to faint hybridization signals. *c*, Schematic summary of the chromosome 2 polymorphisms observed, representing the major population within a parasite line. Polymorphisms revealed within the isolates ABC and 7G8 are indicated by arrows. STS probes numbered from left to right correspond to those in Table 1. The locations of these STS probes are indicated by vertical lines, with a margin of error of 3%. *Sma*I cleavage sites are denoted by solid triangles. Broken lines indicate regions of the chromosome that contain either multiple internal deletions of STS markers or are not of FCR3-type origin. The geographic origins of the *P. falciparum* strains investigated are: FCR3 and D3, Gambia; D10, Papua New Guinea; IT2 and 7G8, Brazil; T26 and ABC, Tanzania; HB2, Honduras; FVO⁺, Vietnam; Dd2, Indochina; K1, Thailand.

METHODS. Parasites were cultured under standard conditions²⁴. Parasite DNA embedded in agarose plugs⁹ was separated by pulse-field gradient electrophoresis using a LKB Pulsaphor as follows: *a*, ramped pulse time from 25 to 60 s over 36 h at 160 V, 1% LE agarose, 0.5 × TBE, at 14 °C; *b*, ramped pulse time from 60 to 300 s over 48 h at 150 V, 1% LE agarose, 0.5 × TBE, at 14 °C.



indicate that a complete, representative contig was obtained spanning the entire 1.03 Mb (10^6 bases) of chromosome 2 from telomere to telomere.

Chromosome 2 is subject to breakage and telomere addition within the *KAHRP* gene, resulting in *KAHRP*⁻ parasites and a truncated chromosome¹². However, both *KAHRP*⁺ and *KAHRP*⁻ parasites show chromosome-2 polymorphisms, suggesting that polymorphisms in chromosome 2 can be generated by mechanisms other than breakage and healing within the *KAHRP* gene. The STS defined for chromosome 2 were therefore used to analyse the polymorphisms associated with this chromosome. Genomic DNAs from independent geographic isolates were compared with strain FCR3 by digestion with the restriction endonuclease *Sma*I, generating an internal 235-kilobase (kb) fragment and two telomere-containing fragments of 510 kb and 280 kb. Upon Southern transfer, the nitrocellulose filter was sequentially hybridized with STS 7, 21 and 28. Whereas the central 235-kb *Sma*I fragment is conserved among the strains tested, both terminal *Sma*I fragments are polymorphic (Fig. 2a). Interestingly, these polymorphisms extend to subpopulations within initially clonal lines, indicating a continuing process of chromosomal variation. To define the structure of these polymorphisms more precisely, an expanded series of STS probes were used to probe electrophoretically separated intact chromosomes from these strains. The central 800-kb region of chromosome 2 is conserved (Fig. 2b) (STS 5 to 28 hybridize to all the strains and subpopulations tested), whereas STS probes derived from the subtelomeric regions are polymorphic and do not hybridize to all strains. Several types of subtelomeric polymorphisms emerge from these studies: (1) breakage and healing by telomere addition at the *KAHRP* locus (designated as the left end of the chromosome) account for the variations observed in strains FVO⁻, D3 and Dd2, and in subpopulations of 7G8 (refs 12, 13); (2) subtelomeric deletions at the right end, as in strain ABC and in subpopulations of all the strains investigated, compatible with a breakage and healing type of mechanism; (3) internal deletions or substitution of DNA not of FCR3-chromosome 2-type origin, as in strain K1 (80 kb at the left end) and HB2 (90 kb at the right end); (4) insertion or duplication of DNA within the right subtelomeric region of strain Dd2. In all

cases the polymorphisms are constrained within a 120-kb region at either end of the chromosome, thereby defining a subtelomeric region of structural instability.

Chromosomal polymorphisms in other eukaryotic systems are frequently associated with regions of unusual transcriptional activity¹⁴⁻¹⁶. We therefore sought to determine whether differences could be detected between the transcription pattern of the polymorphic and non-polymorphic regions of the chromosome. RNA was isolated from asynchronous, intraerythrocytic-stage cultures of FCR3, fractionated into poly(A)⁺ and poly(A)⁻ populations and labelled by reverse transcriptase/PCR amplification¹⁷. These complementary DNA probes were hybridized to the ordered array of YAC clones (Fig. 3). Poly(A)⁺ RNA-derived probes hybridized only to YAC clones spanning the conserved 80% of the chromosome. In contrast, neither poly(A)⁺- nor poly(A)⁻ RNA-derived probes hybridized to the polymorphic ends of the chromosome. However, poly(A)⁻ RNA-derived probes hybridize to the YAC clone HG12 (Fig. 3b). Thus, within the limits of resolution provided by these YAC clones and the sensitivity of the transcription mapping technique¹⁷, differences in transcription are observed along the chromosome. These data suggest that chromosome 2 is compartmentalized during the intraerythrocytic stages into a conserved central domain encoding poly(A)⁺ transcripts and transcriptionally silent polymorphic ends. Transcription mapping of other random telomere YAC clones supports the model that intraerythrocytic-stage genes are excluded from subtelomeric regions. But in the region of frequent breakage and healing events at the left end of the chromosome, we found transcripts for both poly(A)⁺ and poly(A)⁻ RNA. We speculate that the clustering of these transcripts may play a role in the generation of subtelomeric instability at this end of chromosome 2.

The structural basis for the chromosomal polymorphisms of chromosome 2 appears to result from rearrangement and deletion events within the subtelomeric regions, a feature likely to be common to many of the 14 chromosomes of *P. falciparum*^{7,8}. The appearance of subpopulations in initially clonal parasite lines suggests that chromosomal alterations in subtelomeric regions occur frequently and continuously during mitotic division. These chromosomal polymorphisms are observed not

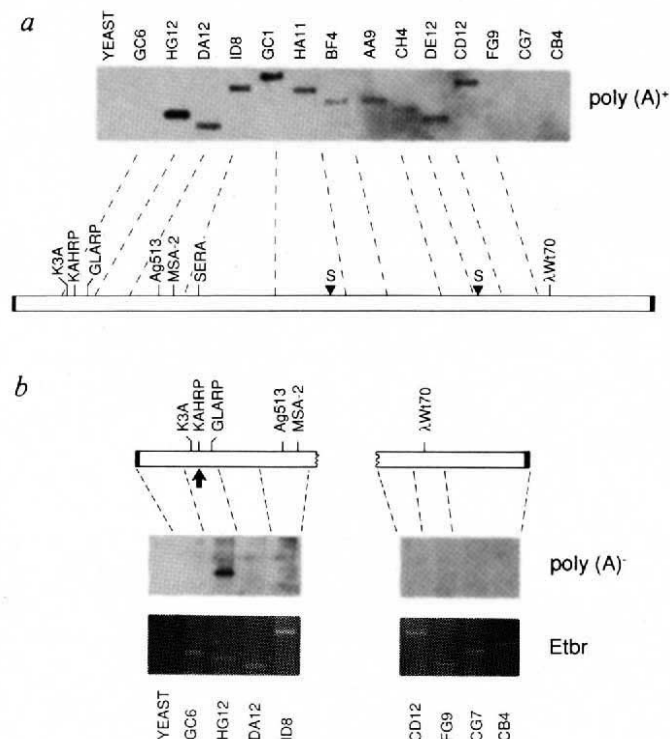


FIG. 3 Transcription mapping of *P. falciparum* chromosome 2. Southern blots containing an ordered array of the minimal representative YAC clones for chromosome 2 (see Fig. 1a for comparison) were hybridized with labelled cDNA probes generated by reverse transcription and amplification of poly(A)⁺ RNA (a) or poly(A)⁻ RNA (b). RNA was prepared from an asynchronously growing intraerythrocytic culture of the *P. falciparum* strain FCR3 and incubated with 0.1 units per μ l of DNase I for 30 min at 37 °C. The ethidium bromide-stained pulse-field gel (Etbr) containing representative YAC clones is shown. No hybridization of poly(A)⁻ RNA-derived probes to internal YAC clones was observed. An arrow indicates a region of the chromosome that is subject to frequent breakage and healing events. Broken lines reveal the chromosomal origin of the hybridization signals observed. The positions of relevant markers and restriction sites (S, *Sma*I) along chromosome 2 are indicated.

METHODS. The conditions for pulse-field gel electrophoresis are described in the legend to Fig. 1b. For transcription mapping of artificial chromosomes 1 μ g poly(A)⁺ RNA or 10 μ g poly(A)⁻ RNA was reverse-transcribed into cDNA using random primers¹⁷. The cDNA was amplified in the presence of [α -³²P]dCTP using the TAG-IT kit (BIOS) which uses d(N)₆(GC)(GC)(GC) as primers. The amplified labelled cDNA was purified by column chromatography to remove oligonucleotides smaller than 100 bp.

only in cultured parasites but also in field isolates^{18,19}, suggesting a potential biological function for these events in the survival of the parasite. □

Received 1 October; accepted 3 December 1992.

- Wellems, T. E. *et al. Cell* **49**, 633–642 (1987).
- Peterson, D. S., Walliker, T. E. & Wellems, T. E. *Proc. natn. Acad. Sci. U.S.A.* **85**, 9115–9118 (1988).
- Wellems, T. E. *et al. Nature* **345**, 253–255 (1990).
- Walliker, D. *et al. Science* **236**, 1661–1666 (1987).
- Goman, M. *et al. Molec. Biochem. Parasit.* **5**, 391–400 (1982).
- Pollack, Y., Katzen, A. L. & Golenser, J. *Nucleic Acids Res.* **10**, 539–546 (1982).
- Kemp, D. J. *et al. Nature* **315**, 347–350 (1985).
- van der Ploeg, L. H. T. *et al. Science* **229**, 658–661 (1985).
- de Bruin, D., Lanzer, M. & Ravetch, J. V. *Genomics* **14**, 332–339 (1992).
- Walker-Jonah, A., Dolan, S. A., Gwadz, R. W., Panton, L. J. & Wellems, T. E. *Molec. Biochem. Parasit.* **51**, 313–320 (1992).
- Patarapotikul, J. & Langsley, G. *Nucleic Acids Res.* **16**, 4331–4340 (1988).

- Pologe, L. G. & Ravetch, J. V. *Nature* **322**, 474–477 (1986).
- Pologe, L. G. & Ravetch, J. V. *Cell* **55**, 869–874 (1988).
- Lutzer, S., Rothman, P., Pollock, R., Coffman, R. & Alt, F. W. *Cell* **53**, 177–184 (1988).
- Blackwell, T. K. & Alt, F. W. *A. Rev. Genet.* **23**, 605–636 (1989).
- Borst, P. & Greaves, D. R. *Science* **235**, 658–667 (1987).
- Lanzer, M., de Bruin, D. & Ravetch, J. V. *EMBO J.* **11**, 1949–1955 (1992).
- Corcoran, L. M., Forsyth, K. P., Bianco, A. E., Brown, G. V. & Kemp, D. J. *Cell* **44**, 87–95 (1986).
- Biggs, B. A., Kemp, D. J. & Brown, G. V. *Proc. natn. Acad. Sci. U.S.A.* **86**, 2428–2432 (1989).
- Triglia, T. *et al. EMBO J.* **6**, 1413–1419 (1987).
- Smythe, J. A. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 5195–5199 (1988).
- Fenton, B. *et al. Molec. cell. Biol.* **11**, 963–971 (1991).
- Bzik, D. J., Wurbo, L., Horii, T. & Inselburg, J. *Molec. Biochem. Parasit.* **30**, 279–288 (1988).
- Trager, W. T. & Jansen, J. B. *Science* **193**, 673–675 (1976).

ACKNOWLEDGEMENTS. We thank P. Model and H. Stunnenberg for critically reading the manuscript, S. Wertheimer for discussion, C. Ralphe and M. Samuels for technical assistance, and C. Ritter for secretarial support. This work was supported by the World Health Organization and a grant from the US army. M.L. is supported by fellowships from the Deutscher Akademischer Austauschdienst and from the World Health Organization.

Two nuclear signalling pathways for vitamin D

Carsten Carlberg*, Igor Bendik, Adrian Wyss, Elke Meier, Laurie J. Sturzenbecker†, Joseph F. Grippo† & Willi Hunziker

Department of Pharma Research New Technologies, F. Hoffmann–La Roche Ltd, CH-4002 Basel, Switzerland

† Department of Toxicology and Pathology, Hoffmann–La Roche, Nutley, New Jersey 07110, USA

THE dihydroxylated form of vitamin D₃ (1,25-dihydroxy-D₃) mediates a biological response by binding to intracellular receptors^{1–4} which belong to the steroid receptor superfamily⁵. These receptors act as ligand-dependent transcription factors that bind to specific DNA sequences (reviewed in refs 6–9). We have identified two classes of vitamin D response elements that are activated either by the vitamin D receptor (VDR) alone or by heterodimers of VDR and the retinoid-X receptor- α (RXR- α)¹⁰. The motif GGGTGA arranged as a direct repeat with a spacing of six nucleotides or as a palindrome without spacing, or as an inverted palindrome with a 12-nucleotide spacing, confers vitamin D inducibility mediated by VDR alone. A second class of response elements, composed of directly repeated pairs of motifs (GGTCCA, AGGTCA, or GGGTGA) spaced by three nucleotides, is synergistically activated by RXR and VDR, but only in the presence of both ligands. Thus, the RXR ligand and the nature of the response element determine whether a nuclear receptor is co-regulated by RXR.

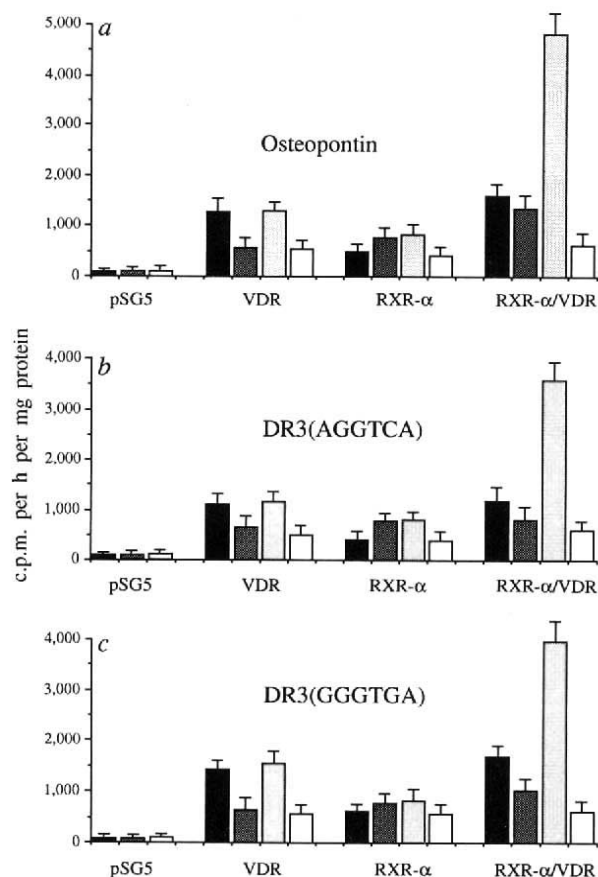
Recent evidence indicates that the retinoid/thyroid hormone/vitamin D subfamily of nuclear receptors might bind to their response elements only as heterodimers with RXR (refs 11–16). We wondered if both ligands, 1,25-dihydroxy-vitamin D₃ (referred to as vitamin D) and 9-*cis*-retinoic acid^{17,18}, are required for a VDR/RXR heterodimer-mediated transcriptional response and whether heterodimerization with RXR is a prerequisite for a vitamin D effect. *Drosophila* SL-3 cells are a good cellular system to answer this question, as they are devoid of VDR and RXR. The cells respond to the ligands upon transfection of the respective receptor expression plasmids together with an appropriate response element driving the expression of the chloramphenicol acetyltransferase (CAT) reporter gene (Fig. 1). A single copy of the mouse osteopontin vitamin D response element, a perfect direct repeat of the motif GGTCA spaced by three nucleotides¹⁹, and single copies of the direct repeats of the motifs AGGTCA²⁰ and GGGTGA spaced by three nucleotides, conferred a ~2-fold CAT induction by vitamin D or 9-*cis*-retinoic acid (the ligand for RXR^{17,18}) when the corres-

ponding receptor expression plasmid was cotransfected. Transfection of both receptor expression plasmids and stimulation with individual ligands induced transcription to levels similar to those observed when only the cognate receptor expression plasmid was transfected. Stimulation with both ligands, however, showed a clear synergistic response. Thus, the data show that this class of vitamin D response elements (VD/RXREs) is preferentially activated by VDR/RXR heterodimers. This notion is further supported by band-shift assays using *in vitro*-translated receptors and the osteopontin response element as a probe. A prominent shifted complex is observed in the presence of both receptors (Fig. 2, and ref. 13). VDR alone or with SL-3 cell extract shows no complex, indicating that neither the reticulocyte lysate nor the SL-3 cell extract contains an RXR-like activity. Thus, the affinity of VDR for this element *in vitro* is greatly enhanced in the presence of RXR- α . The binding is, however, ligand-independent (data not shown). In contrast, the synergistic effect of the two receptors on transcription requires the presence of the ligands for both receptors.

We next examined the properties of a second set of vitamin D response elements, those of the human osteocalcin gene^{21–24} and some variants thereof that emerged from an extensive mutational analysis of this region (data not shown). The osteocalcin vitamin D response region, or two copies of the motif PuGGTGA arranged as a direct repeat spaced by six nucleotides, as a palindrome without spacing, or as an inverted palindrome spaced by 12 nucleotides, was found to confer vitamin D inducibility to the CAT reporter gene when cotransfected with the VDR expression plasmid. Cotransfection of RXR and stimulation with 9-*cis*-retinoic acid showed no effect. Surprisingly, and in contrast to the class of elements analysed above, no synergistic effect on transcription was observed in the presence of both receptors and both ligands (Fig. 3). Interestingly, similarly spaced arrangements of the motifs PuGTTCa or PuGGTCA showed no inducibility by vitamin D or 9-*cis*-retinoic acid (data not shown). The data suggest that this class of vitamin D response elements (VD-REs) is activated independently of RXR, possibly by VDR homodimers. The RXR-independence of the response is supported by band-shift analysis using the osteocalcin VD-RE region as a probe. VDR translated *in vitro* (or RXR) shows a weak shifted complex which is not enhanced by the addition of RXR- α or SL-3 cell extract (Fig. 2). Thus, in contrast to the results with the osteopontin element, RXR- α does not increase the affinity of VDR for the osteocalcin element. The similar migration but dissimilar intensities of these two shifted complexes suggests a lower-affinity binding of a VDR homodimer to the osteocalcin element. The discrepancy between the transactivation potency of the two classes of elements and their ability to bind *in vitro*-translated receptor with high affinity suggests that *in vivo* the occupied receptor might be modified²⁵.

The lack of interference between the two transfected receptor systems in *Drosophila* cells encouraged us to address the same

* Present address: Clinique de Dermatologie, Hôpital Cantonal Universitaire, CH-1211 Genève, Switzerland.



question in mammalian cells which contain endogenous VDR and RXR- α . MCF-7 cells transfected with the reporter plasmid containing the osteopontin or the other VD/RX-REs showed only a marginal response to vitamin D or 9-*cis*-retinoic acid, whereas the presence of both ligands resulted in a dramatic induction of CAT expression (Fig. 4). The synergy was even more pronounced than that observed in SL-3 cells (Fig. 1). In

FIG. 2 Band shift analysis of the binding of *in vitro* translated VDR and RXR- α to the osteopontin and the osteocalcin elements. With the osteopontin probe an intense shifted complex (arrow) is observed only in the presence of both receptors, indicating the binding of VDR/RXR heterodimers to this element. In contrast, with the osteocalcin element only a weak shifted complex is observed in the presence of VDR (Δ); this is not enhanced by the addition of RXR or *Drosophila* SL-3 cell extract (which shifts the osteocalcin element to a different position; diamond symbol).

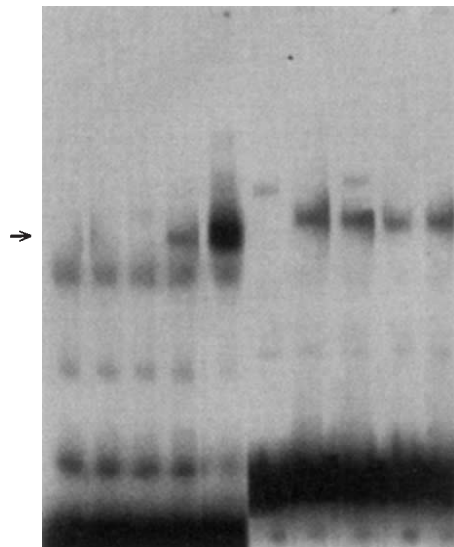
METHODS. The complementary DNAs of the human VDR and the human RXR- α cloned in pSG5 were linearized by *Xba*I and *Bgl*II, respectively, for *in vitro* transcription under standard conditions. For *in vitro* translation, 5 μ g RNA were incubated (30 °C for 90 min) with 175 μ l rabbit reticulocyte lysate (Promega), 100 U RNasin (Promega), and a complete amino-acid mixture (20 μ M) in a total volume of 250 μ l. All response elements were isolated by restriction digestion (with *Hind*III and *Bam*HI) of the appropriate response element-containing pBLCAT2 plasmids and were labelled by a fill-in reaction using [α - 32 P]dCTP and T7 DNA polymerase (USB). For the binding reaction, 5 μ l of the individual *in vitro*-translated receptors (or 2.5 μ l each, for the mixture of the two) and 1 μ l *Drosophila* SL-3 cell extract (prepared essentially as described in ref. 32, and added where indicated) were preincubated with the appropriate ligand (1 μ M) for 10 min at room temperature in a total volume of 20 μ l binding buffer (10 mM HEPES (pH 7.9), 40 mM KCl, 1 mM DTT, 0.2 μ g μ l $^{-1}$ BSA, 0.2 μ g μ l $^{-1}$ poly(dI-dC) and 5% glycerol), before addition of the probe (~1 ng, corresponding to 25,000 c.p.m.) and further incubation for 30 min at room temperature. The protein-DNA complexes were resolved on a 5% non-denaturing polyacrylamide (29:1) gel at room temperature in 0.5 $\times$ TBE (pre-run for 2 h), then the dried gel was exposed at -80 °C to a Kodak XAR film with an intensifying screen.

FIG. 1 Heterodimerization of VDR and RXR- α on the mouse osteopontin VD/RX-RE and on other directly repeated hexamer motifs spaced by 3 nucleotides. *Drosophila* SL-3 cells were cotransfected with pBLCAT2 vectors³⁰ containing the mouse osteopontin VD/RX-RE (a direct repeat of the motif GGTCA)¹⁹ (a) or direct repeats of the motifs AGGTCA (b), and GGTGA (c), and the indicated receptor expression vectors VDR, RXR- α , both (RXR- α /VDR) or the expression vector without insert (pSG5). Cells were treated for 40 h with 1 μ M vitamin D (black bars), 1 μ M 9-*cis*-retinoic acid (dark grey bars), 1 μ M vitamin D plus 1 μ M 9-*cis*-retinoic acid (light grey) or vehicles (white). Without receptor co-transfection, little CAT activity is observed, indicating the absence of RXR and VDR in SL-3 cells. A synergistic effect observed in the presence of both receptors and both ligands probably reflects the activity of VDR/RXR heterodimers.

METHODS. The response elements used were cloned as double-stranded oligonucleotides with *Xba*I and *Spe*I overhangs into the *Xba*I site upstream of the tk promoter in the CAT reporter plasmid pBLCAT2 (ref. 30). The sequences of the top strands were 5'-ctagtGCTCGGGTACGGTTCACGAGGTTCACTCGACTCGT-3' for the osteopontin element, 5'-ctagtGCTCGGGTACGGTTCACGAGGTTCACTCGACTCGT-3' for DR3(AGGTCA) and 5'-ctagtGCTCGGGTACGGTTCACGAGGTTCACTCGACTCGT-3' for DR3(GGGTGA). SL-3 cells (2×10^6 per 60-mm Petri dish) were grown overnight in Schneider's medium supplemented with 15% charcoal-treated fetal calf serum (FCS). For transfection, liposomes were formed by incubating 5 μ g reporter plasmid, 3 μ g reference plasmid (pRSV β -gal) and 1 μ g of the VDR or RXR- α expression vectors with 30 μ g DOTAP (*N*-(1-(2,3-dioleoyloxy)propyl)-*N,N,N*-trimethylammonium methyl sulphate; Boehringer-Mannheim) for 15 min at room temperature in a total volume of 100 μ l. After dilution with 0.9 ml Schneider's medium, the liposomes were added to the cells. Vitamin D, 9-*cis*-retinoic acid (final concentration, 1 μ M each) or solvent (ethanol) was added 8 h after transfection. Cells were collected 40 h after the start of treatment for determination of CAT activity³¹.

contrast, transfection of MCF-7 cells with the VD-REs yielded the same level of transactivation, regardless of whether cells were treated with vitamin D alone or in combination with 9-*cis*-retinoic acid (Fig. 4). The absolute CAT expression levels obtained with the osteocalcin VD-RE region were ~3-fold higher than those with the other VD-REs; this may be due to the presence of an AP-1 site in this region.

Probe	Osteopontin				Osteocalcin			
RXR- α	-	-	-	+	-	-	-	+
VDR	-	+	+	-	-	+	+	-
SL-3	+	-	+	-	+	-	+	-



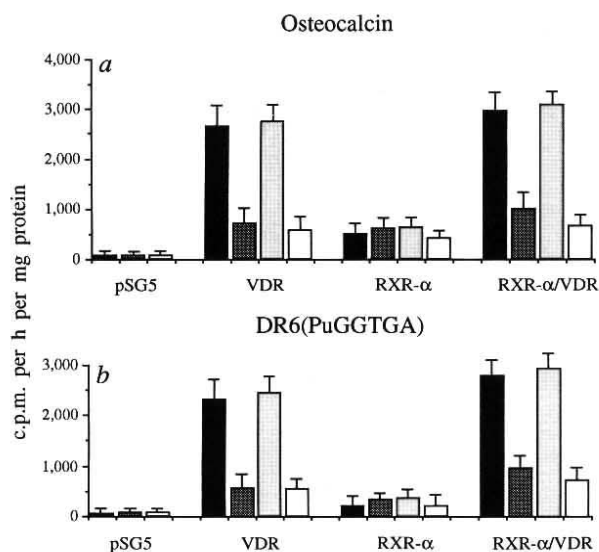
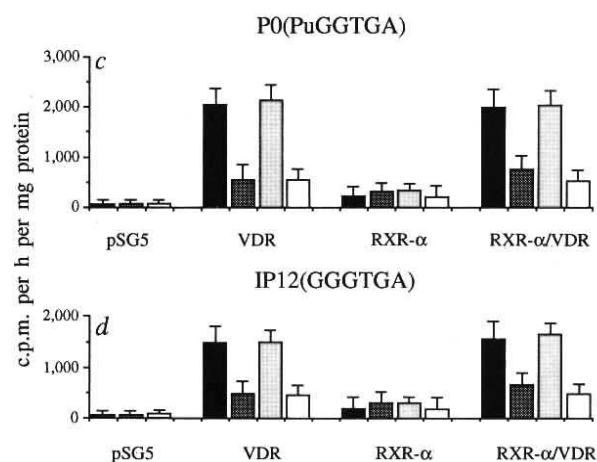


FIG. 3 RXR-independent activity of VDR on the human osteocalcin VD-RE and on direct repeats, palindromes or inverted palindromes of the motif PuGGTGA with spacing of 6, 0 and 12 nucleotides, respectively. *Drosophila* SL-3 cells were cotransfected with pBLCAT2 vectors containing the human osteocalcin vitamin D response region (a), or the response motif PuGGTGA either as a direct repeat spaced by 6 nucleotides (DR6(PuGGTGA); b, as a palindrome without spacing (PO(PuGGTGA); c, as an inverted palindrome with 12 nucleotides spacing (IP12(GGGTGA); d, or either the pSG5 expression vector without insert (pSG5) or with VDR, RXR- α , or both receptor expression vectors (RXR- α /VDR). Cells were treated for 40 h with 1 μ M vitamin D (black bars), 1 μ M 9-*cis*-retinoic acid (dark grey), 1 μ M vitamin D plus 1 μ M 9-*cis*-retinoic acid (light grey), or vehicle (white). All response elements responded to vitamin D, but no synergy was observed in the presence of



both receptors and both ligands. These data indicate that VDR activates these elements independently of RXR, possibly as a homodimer. METHODS. The response elements used were cloned as described in the legend to Fig. 1. The sequences of the top strands were: 5'-ctagtGCTCGGGTAGGGGTGACTACCGGGTGAACGGGGGCACTCGACTCGT-3' for the human osteocalcin vitamin D response region, 5'-ctagtGCTCGGGTAGGGGTGACCGCAAAGGTGACTCGACTCGT-3' for DR6(PuGGTGA), 5'-ctagtGCTCGGGTAGGGGTGATCACCTCGACTCGT-3' for PO(PuGGTGA), and 5'-ctagtGCTCGGGTAGGGGTGATCACCTCGACTCGT-3' for IP12(GGGTGA). The PuGGTGA motifs are shown in bold face and the *Xba*I and *Spe*I overhangs are shown in lower case. Transfection of the cells and analysis was done as described in the legend to Fig. 1.

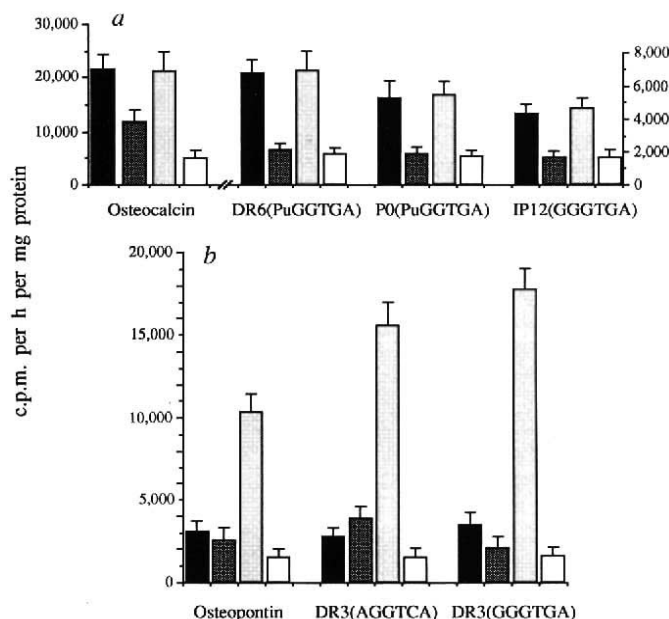
In both *Drosophila* and mammalian cells, neither class of elements is influenced by RXR in the absence of its ligand. In the presence of the ligand, RXR did not quench the vitamin D response mediated by VD-REs, even in MCF-7 cells containing physiologically low receptor levels. This suggests that the ternary complex between the VD-RE and a homodimer of liganded VDR is more stable than the heterodimeric complex between liganded VDR and RXR in solution. Furthermore, the similar vitamin D response obtained in SL-3 cells from VD/RX-REs

in the presence and absence of RXR- α cotransfection suggests the binding of VDR homodimers to these elements in the absence of 9-*cis*-retinoic acid. This would imply that, *in vivo*, the formation of the ternary complex between response element and receptor dimer is ligand-dependent. This concept is supported by a recent report that the binding of RXR homodimers to response elements is ligand-dependent²⁶.

Taken collectively, these data demonstrate the existence of two functionally distinct classes of response elements for vitamin

FIG. 4 RXR-dependent and -independent vitamin D responses mediated by endogenous receptors in MCF-7 cells. Cells were transfected with the pBLCAT2 reporter plasmids containing the indicated response elements. The cells were treated for 40 h with 1 μ M vitamin D (black bars), 1 μ M 9-*cis*-retinoic acid (dark grey), 1 μ M vitamin D plus 1 μ M 9-*cis*-retinoic acid (light grey) or vehicle (white) alone. The osteocalcin vitamin D response region and the other three elements activated by VDR in an RXR-independent fashion (Fig. 2) responded to vitamin D, but no synergy was observed in the presence of both ligands (a). However, a synergistic effect of the two ligands was observed with the osteopontin VD/RX-RE and the two other direct repeats spaced by 3 nucleotides (b).

METHODS. Reporter plasmids are described in the legends to Figs 1 and 2. The human breast cancer cell line MCF-7 was grown in RPMI 1640 supplemented with 10% FCS. This cell line expresses VDR and RXR- α (not shown). For transfection, cells (5×10^5 per 60-mm Petri dish) were grown overnight in phenol-red-free RPMI 1640 supplemented with 10% charcoal-treated FCS. Liposomes were formed by incubating 5 μ g reporter plasmid and 3 μ g reference plasmid pRSC β -gal with 30 μ g DOTAP (Boehringer-Mannheim) for 15 min at room temperature in a total volume of 100 μ l. After dilution with 0.9 ml phenol-red-free RPMI 1640, the liposomes were added to the cells. Treatment and analysis of the cells are described in the legend to Fig. 1.



D: one activated by vitamin D alone, possibly through VDR homodimers, and the other activated by vitamin D and 9-*cis*-retinoic acid VDR/RXR heterodimers. The fact that two natural vitamin D response elements, namely that of mouse osteopontin gene and that of the human osteocalcin gene, are activated differentially, one in an RXR-dependent and the other in an RXR-independent fashion, supports the physiological significance of the two nuclear signalling pathways for vitamin D.

Although it is accepted that RXRs can form transcriptionally

active heterodimers with vitamin D, retinoid and thyroid hormone receptors¹¹⁻¹⁶, it remains to be seen whether functioning as an accessory factor for these receptors²⁷⁻²⁹ is a property exclusive to RXRs, or if other receptors can function in the same manner. In the first case a dual signalling pathway may also be predicted for the other members of this receptor subgroup, as response elements seem capable of discriminating between different dimers. The latter possibility, however, would imply that the combinatorial complexity of signalling pathways for these hormones might be even greater than anticipated. □

Received 25 August; accepted 4 December 1992.

- Minghetti, P. P. & Norman, A. W. *FASEB J.* **2**, 3043-3053 (1988).
- Pike, J. W. *A. Rev. Nutr.* **11**, 189-216 (1991).
- DeLuca, H. F., Krisinger, J. & Darwish, H. *Kidney Int.* **38**, S2-S8 (1990).
- Hausler, M. R. *et al. Recent prog. Horm. Res.* **44**, 263-297 (1988).
- Baker, A. R. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 3294-3298 (1988).
- Green, S. & Chambon, P. *Trends Genet.* **4**, 309-314 (1988).
- Evans, R. M. *Science* **240**, 889-895 (1988).
- Beato, M. *Cell* **56**, 335-344 (1989).
- O'Malley, B. W. *Molec. Endocr.* **4**, 363-369 (1990).
- Mangelsdorf, D. J., Ong, E. S., Dyk, J. A. & Evans, R. M. *Nature* **345**, 224-229 (1990).
- Yu, V. C. *et al. Cell* **67**, 1251-1266 (1991).
- Zhang, X. K., Hoffmann, B., Trans, P. B. V., Graupner, G. & Pfahl, M. *Nature* **355**, 441-446 (1992).
- Kliwer, S. A., Umesono, K., Mangelsdorf, D. J. & Evans, R. M. *Nature* **355**, 446-449 (1992).
- Leid, M. *et al. Cell* **68**, 377-395 (1992).
- Bugge, T. H., Pohl, J., Lonnoy, O. & Stunnenberg, H. G. *EMBO J.* **11**, 1409-1418 (1992).
- Marks, M. S. *et al. EMBO J.* **11**, 1419-1435 (1992).
- Levin, A. A., *et al. Nature* **355**, 359-361 (1992).
- Heyman, R. A. *et al. Cell* **68**, 397-406 (1992).

- Noda, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 9995-9999 (1990).
- Umesono, K., Murakami, K., Thompson, C. C. & Evans, R. M. *Cell* **65**, 1255-1266 (1991).
- Morrison, N. A. *et al. Science* **246**, 1158-1161 (1989).
- Kerner, S. A., Scott, R. A. & Pike, J. W. *Proc. natn. Acad. Sci. U.S.A.* **86**, 4455-4459 (1989).
- Schüle, R. *et al. Cell* **61**, 497-504 (1990).
- Ozono, K., Liao, J., Kerner, S. A., Scott, R. A. & Pike, J. W. *J. Biol. Chem.* **265**, 21881-21888 (1990).
- Bagchi, M. K., Tsai, S. Y., Tsai, M.-J. & O'Malley, B. W. *Proc. natn. Acad. Sci. U.S.A.* **89**, 2664-2668 (1992).
- Zhang, X.-K. *et al. Nature* **358**, 587-591 (1992).
- Liao, J., Ozono, K., Sone, T., McDonnell, D. P. & Pike, J. W. *Proc. natn. Acad. Sci. U.S.A.* **87**, 9751-9755 (1990).
- Sone, T., Ozono, K. & Pike, J. W. *Molec. Endocr.* **5**, 1578-1586 (1991).
- Ross, T. K., Moss, V. E., Pahl, J. M. & DeLuca, H. F. *Proc. natn. Acad. Sci. U.S.A.* **89**, 256-260 (1992).
- Luckow, B. & Schütz, G. *Nucleic Acids Res.* **15**, 5490 (1987).
- Gorman, C. M., Moffat, L. F. & Howard, B. H. *Molec. cell. Biol.* **2**, 1044-1051 (1982).
- Schreiber, E., Matthias, P., Müller, M. M. & Schaffner, W. *Nucleic Acids Res.* **17**, 6419 (1989).

ACKNOWLEDGEMENTS. We thank M. Klaus for 9-*cis*-retinoic acid, G. Schütz for the pBLCAT plasmids and for discussion, and W. Wahli for *Drosophila* Si-3 cells.

Evidence for an essential non-Watson-Crick interaction between the first and last nucleotides of a nuclear pre-mRNA intron

Roy Parker* & Paul G. Siliciano†

*Department of Molecular and Cellular Biology, Life Sciences South, University of Arizona, Tucson, Arizona 85721, USA

†Department of Biochemistry and Institute of Human Genetics, University of Minnesota Medical School, 4-225 Millard Hall, 435 Delaware Street South East, Minneapolis, Minnesota 55455, USA

NUCLEAR pre-messenger RNA splicing requires the action of five small nuclear (sn) RNAs, U1, U2, U4, U5 and U6, and more than 50 proteins^{1,2}. The mechanistic similarity of nuclear pre-mRNA splicing and group II self-splicing suggests that many of the central processes of nuclear pre-mRNA splicing are based on RNA-RNA interactions^{3,4}. To understand the mechanism of pre-mRNA splicing, the interactions, and their temporal relationships, that occur between the snRNAs and the pre-mRNA during splicing must be identified. Several snRNA-snRNA⁵⁻⁹ and snRNA-intron⁹⁻¹² interactions have been demonstrated but the putative RNA-based interactions that recognize the AG dinucleotide at the 3' splice site during 3' cleavage and exon ligation are unknown. We report here the reciprocal suppression between 5' and 3' splice site mutations in the yeast actin intron, and propose that the 3' splice site is positioned for 3' cleavage and exon ligation, at least in part, through a non-Watson-Crick interaction between the guanosines at the 5' and 3' splice sites.

After the first step of splicing (5' cleavage and lariat formation), the choice of the 3' splice site depends on the AG dinucleotide both in yeast and in mammals¹³⁻¹⁵. Point mutations in the AG of the yeast *ACT1* intron (U302, C302, A303, U303 and

C303 (Fig. 1a)) inhibit the second step in splicing (3' cleavage and exon ligation) (Fig. 1b; refs 13, 16). The 5' splice site sequences also play a role in the second step of splicing. For example, mutation of the highly conserved GU at the 5' splice site (for example GU→AU, termed the A1 mutation, Fig. 1) blocks the second step causing accumulation of the lariat intermediate and preventing the production of mature mRNA (Table 1, Fig. 1b; refs 13, 17-19).

These observations, and the importance of non-Watson-Crick interactions in specifying RNA interactions²⁰⁻²³, led us to hypothesize that the AG dinucleotide at the 3' splice site would interact with the GU of the 5' splice site in a non-Watson-Crick manner during the second step of splicing. We examined splicing in *Saccharomyces cerevisiae* by primer extension analysis of an *ACT1-CUP1* fusion transcript containing various mutations at the 5' and/or 3' splice sites. In addition, as an indirect but sensitive assay for mature mRNA, we examined the effects of these mutations on the production of β -galactosidase activity from an *ACT1-lacZ* fusion construct bearing the same point mutations (Table 1).

One prediction of this hypothesis is that defects in the second step of splicing caused by a mutation at the 5' splice site could be suppressed by alteration of the 3' splice site in a manner that restores this interaction. In this light, we examined the splicing of introns containing mutations at the 5' and/or 3' splice sites. A critical observation is that the defect in 3' cleavage and exon ligation caused by the A1 mutation can be suppressed by alteration of the 3' splice site. For example, combination of the A1 and C303 mutations restores mature mRNA levels to about 10% of the wild-type levels as judged by lacZ assays of *ACT1-lacZ* fusion constructs (Table 1) and by primer extension analysis of the *ACT1-CUP1* mRNA (Fig. 1b). Moreover, this suppression is reciprocal; the C303 mutation is also suppressed by the A1 mutation (Table 1, Fig. 1b).

Other mutations in the AG dinucleotide have varying effects on the A1 mutation. The A303 mutation weakly suppresses the A1 lesion and mature mRNA accumulates to 3% of wild-type levels (Table 1 and Fig. 1b). The U303 mutation does allow some processing of lariats containing the A1 lesion but this effect is barely detectable (Table 1 and Fig. 1b). Alteration of

TABLE 1 Activity of lacZ in constructs

Construct	lacZ activity	Construct	lacZ activity
WT	100	A303	2.0
A1	<0.1	A1-U302	<0.1
A1-C303	9.8	A1-C302	<0.1
A1-A303	2.8	U302	2
A1-U303	0.4	C302	4
C303	1.5	UT	<0.1
U303	6.0		

β -galactosidase levels are averages of independent assays on at least three different transformants, and vary typically by less than 10%, done as previously described¹¹. WT, Wild type; UT, untransformed parental strain.

the adenosine of the AG dinucleotide to either uracil (U302) or cytosine (C302) does not allow any suppression of the A1 defect (Table 1 and Fig. 1*b*). These results indicate that suppression is specific both for position and for nucleotide. Consistent with the lacZ levels, the primer extension product found in the A1/C303, A1/A303 and A1/U303 double mutants is exactly the same size as in the wild type (data not shown), indicating that splicing has occurred at the correct 5' and 3' splice sites.

The strongest suppression results in about 10% of the wild-type mature mRNA levels. We believe these levels are significant because in many experiments introns bearing the A1 mutation do not make detectable mature mRNA^{13,17-19}. The suppression is probably limited by several factors. First, a significant portion of the transcripts containing the A1 mutation never complete the first step of splicing, thus reducing the amount of lariat intermediate available for the second step. Second, additional contacts that normally occur with the guanosine bases may be lacking in the suppressed double mutants (for example A1/C303). Similarly, as suggested by the analysis of triple helical interactions in group I introns²¹, the stability of the particular base interactions (such as G1-G303, A1-A303 or A1-C303) may be greatly influenced by neighbouring interactions. A final limi-

tation on the level of suppression arises from the possible involvement of these nucleotides in other substeps in the second step of splicing. Although the addition of some 3' splice site mutations have clearly suppressed a rate-limiting step in the A1 mutant, we do not know whether these substitutions will now limit other steps.

Additional evidence for this interaction is found by sequence comparison of introns with non-consensus 5' and 3' splice sites²⁴. Introns that begin with the non-consensus 5' splice site sequence AU end with the non-consensus 3' splice site AC. Together with the results presented here, this covariation argues that specific combinations of nucleotides at the 5' and 3' splice sites are required for the second step in splicing. Because the observed genetic suppression is allele specific and reciprocal, the interaction is likely to be direct. But our results do not rule out the more complicated possibility of an indirect interaction.

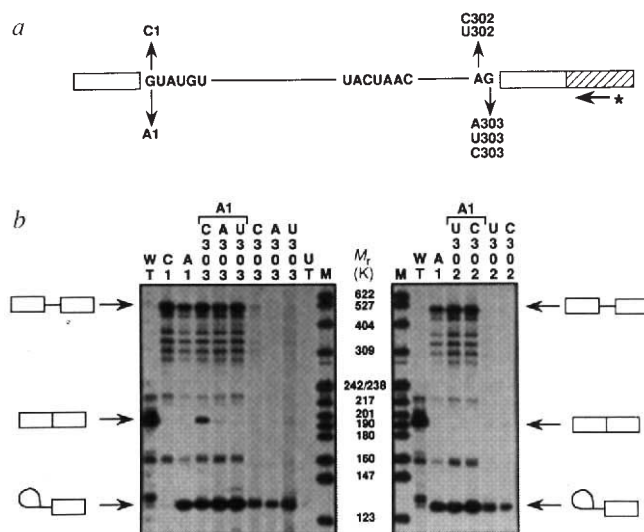
In the simplest model, the proposed interaction would result from one or more hydrogen bonds between the guanosines at the 5' (G1) and 3' (G303) splice sites. Two possible models are shown in Fig. 2a. In both cases analogous interactions can be drawn with A1-C303 and A1-A303. Moreover, similar A1-G303 interactions result in steric overlap, thus providing a possible basis for the step two defect caused by the A1 mutation. To define unambiguously the interaction between G1 and G303, *in vitro* experiments with nucleotide analogues will be necessary.

The proposed G1–G303 interaction may occur at the same time as other RNA-based interactions that could contribute to the positioning of the 3' splice site. For example, other important contacts are likely to be to the A of the AG. Similarly, the hydrogen bonding of the U5 snRNA with the 3' exon is likely to occur at this stage in the reaction¹⁶ (Fig. 2b).

Regardless of the detailed hydrogen-bonding pattern, the data suggest that an important role of the conserved AG dinucleotide is to form RNA-RNA contacts that position the 3' splice site during the second chemical step of splicing. Moreover, although there are clearly some chemical differences in the first and second catalytic steps in splicing²⁵, location of the 3' splice site close to where 5' cleavage occurred raises the possibility, as has been

FIG. 1. *a*, Schematic of the *ACT1* intron in the *CUP1* fusions illustrating the mutations used in this work and the position of the primer used for primer extension. Open boxes represent actin exon sequences, hatched box represents *CUP1* coding sequences and a line is used for intron sequences. Mutations at positions 302 and 303 also include a G→C transversion two nucleotides into the 3' exon (termed C305, ref. 12). This alteration, which has no effect on splicing itself (data not shown) removes an AG dinucleotide that could complicate the analysis if left unaltered. *b*, Primer extension analysis of transcripts produced from the various mutant introns when expressed in the *ACT1-CUP1* fusion. Lanes are as labelled using the following abbreviations; WT, wild-type; UT, untransformed parental strain; M, molecular mass markers. An attempt to extend this analysis using a cytosine at position 1 of the intron, C1, was unsuccessful because we were unable to get a significant portion of the C1 transcript past the first step in splicing even in the presence of a suppressor U1 mutation (data now shown). On longer exposures we detect small amounts of two spliced products from the introns carrying the C303, A303, U303 mutations. One of these products is the correctly spliced mature mRNA, the second product is five nucleotides larger (data not shown), resulting in a mature mRNA in which the downstream reading frame is out of frame. This product presumably arises by the use of UG dinucleotide five nucleotides 5' of the 3' splice site (at position 298), that can compete with a mutated 3' splice site. We do not detect any use of the alternative 3' splice site at position 298 in introns containing the A1 mutation, suggesting that the use of this site requires a G at the 5' splice site, consistent with the G1-G303 interaction proposed in this work.

METHODS. Each mutant intron was inserted by standard techniques into the *ACT1-CUP1* fusion gene²⁸ in the vector pG-1 (ref. 29). In each case the fusion contains 50 nucleotides of the 3' actin exon fused in frame to the *CUP1* gene. These plasmids were introduced into the strain, I4 (relevant genotype; MAT α , Ura3-52, Trp1- Δ 1, cup1 Δ :URA3) by transformation selecting for the plasmid marker *TRP1*. Primer extension was done essentially as described previously¹³ using a primer, 5'-CTTCATTTTGAAGTAAATTAAAT-3'.



located in the *CUP1* coding region. RNA levels were quantified with a Betascope. The reduction in transcript levels in strains with a mutation at position 302 or 303 appears to be the result of a slight dominant growth defect caused by these transcripts when expressed to high levels as in the *CUP1* fusions. As such, these strains tend to maintain the multicopy plasmids at lower copy number. This effect is not seen with the lacZ fusions, which are expressed to a much lower level, or in the double mutant combinations.

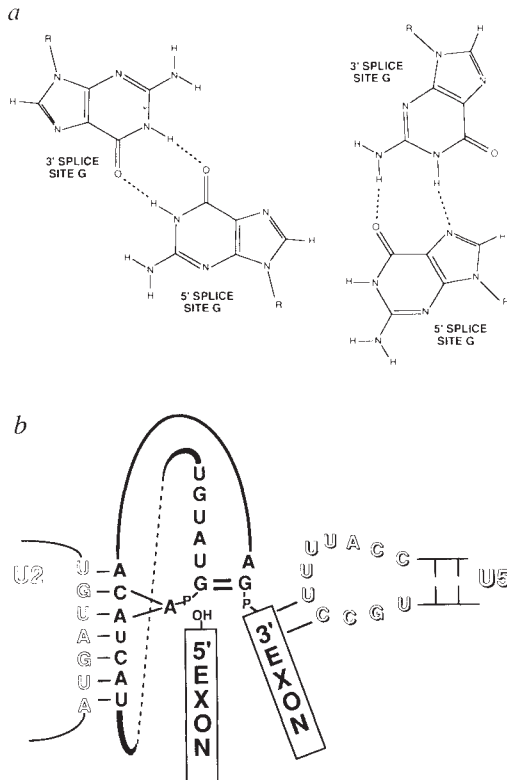


FIG. 2 Two possible models for the interaction between the 5' and 3' guanosine residues of the intron (a) and this potential interaction in the context of other relevant base-pairings (b). a, Note that, with the exception of the A1-A303 possibility in the left model, these base pairs will not be strictly isomorphic with the G1-G303 base pair. This suggests that, as has been seen for an interaction in group I introns between the RNA backbone and a base²³, minor movements of the RNA backbone may be tolerated in this case as well.

suggested for group I introns^{20,26,27}, that essentially the same active site could be used for the second step in splicing as in the first. □

Received 26 August; accepted 1 December 1992.

- Guthrie, C. *Science* **253**, 157-162 (1991).
- Green, M. A. *Rev. Cell Biol.* **7**, 559-599 (1991).
- Guthrie, C. *Am. Zool.* **29**, 557-567 (1989).
- Cech, T. & Bass, B. A. *Rev. Biochem.* **55**, 599-630 (1986).
- Hashimoto, C. & Steitz, J. *Nucleic Acids Res.* **12**, 3283-3293 (1984).
- Brow, D. & Guthrie, C. *Nature* **334**, 213-218 (1988).
- Wu, J. & Manley, J. L. *Nature* **352**, 818-823 (1991).
- Hausner, T., Giglio, L. M. & Weiner, A. W. *Genes Dev.* **4**, 2146-2156 (1990).
- Wasserman, D. & Steitz, J. *Science* **257**, 1918-1925 (1992).
- Zhang, Y. & Weiner, A. *Cell* **46**, 827-835 (1986).
- Parker, R., Siliciano, P. & Guthrie, C. *Cell* **49**, 229-239 (1987).
- Reich, C. I., VanHoy, R. W., Porter, G. L. & Wise, J. A. *Cell* **69**, 1159-1169 (1992).
- Vijayraghavan, U. et al. *EMBO J.* **5**, 1683-1695 (1986).
- Smith, C. W. J., Porro, E. B., Patton, J. G. & Nadal-Ginard, B. *Nature* **342**, 243-247 (1989).
- Reed, R. *Genes Dev.* **3**, 2113-2123 (1989).
- Newman, A. J. & Norman, C. *Cell* **68**, 734-754 (1992).
- Aebi, M., Hornig, H., Padgett, R., Reiser, J. & Weissmann, C. *Cell* **47**, 555-565 (1986).
- Newman, A. J., Ogden, R. C. & Abelson, J. *Cell* **42**, 335-342 (1985).
- Fouser, L. A. & Friesen, J. D. *Cell* **45**, 81-93 (1986).
- Michel, F., Hanna, M., Green, R., Bartel, D. P. & Szostak, J. W. *Nature* **342**, 391-395 (1989).
- Michel, F., Ellington, A. D., Couture, S. & Szostak, J. W. *Nature* **347**, 578-580 (1990).
- Bartel, D. P., Zapp, M. L., Green, M. R. & Szostak, J. W. *Cell* **67**, 529-536 (1991).
- Pyle, A. M., Murphy, F. L. & Cech, T. R. *Nature* **358**, 123-128 (1992).
- Jackson, I. J. *Nucleic Acids Res.* **19**, 3795-3798 (1991).
- Moore, M. J. & Sharp, P. A. *Science* **256**, 992-997 (1992).
- Cech, T. R. *Science* **236**, 1532-1535 (1987).
- Inoue, T., Sullivan, F. X. & Cech, T. R. *J. molec. Biol.* **189**, 143-159 (1986).
- Lesser, C. F. & Guthrie, C. *Genetics* (in the press).
- Schena, M., Picard, D. & Yamamoto, K. R. *Meth. Enzym.* **194**, 389-398 (1991).

ACKNOWLEDGEMENTS. We thank D. Muhlrad for technical assistance and artwork, C. Lesser and C. Guthrie for the *ACT1-CUP1* fusion and A. Adams and D. McPheeters for helpful comments on the manuscript. This work was supported by grants from the Chicago Community Trust to R. P. and the NSF to P. G. S.

PLEASE follow these guidelines so that your manuscript may be handled expeditiously.

Nature is an international journal covering all the sciences. Contributors should therefore bear in mind those readers who work in other fields and those for whom English is a second language, and write clearly and simply, avoiding unnecessary technical terminology. Space in the journal is limited, making competition for publication severe. Brevity is highly valued. One printed page of *Nature*, without interruptions, contains about 1,300 words.

Manuscripts are selected for publication according to editorial assessment of their suitability and reports from independent referees. They can be sent to London or Washington and should be addressed to the Editor. Manuscripts may be dealt with in either office, depending on the subject matter, and will where necessary be sent between offices by overnight courier. High priority cannot be given to presubmission enquiries; in urgent cases they can be made in the form of a one-page fax. All manuscripts are acknowledged on receipt but fewer than half are sent for review. Those that are not reviewed are returned as rapidly as possible so that they may be submitted elsewhere without delay. Contributors may suggest reviewers; limited requests for the exclusion of specific reviewers are usually heeded. Manuscripts are usually sent to two or three reviewers, chosen for their expertise rather than their geographical location. Manuscripts accepted for publication are typeset from the London office.

Nature requests authors to deposit sequence and crystallographic data in the databases that exist for this purpose, and to mention availability of these data.

Once a manuscript is accepted for publication, contributors will receive proofs in about 4 weeks. *Nature*'s staff will edit manuscripts with a view to brevity and clarity, so contributors should check proofs carefully. Manuscripts are generally published 2-3 weeks after receipt of corrected proofs. *Nature* does not exact page charges. Contributors receive a reprint order form with their proofs; reprint orders are processed after the manuscript is published and payment received.

Categories of paper

Review articles survey recent developments in a field. Most are commissioned but suggestions are welcome in the form of a one-page synopsis addressed to the Reviews Coordinator. Length is negotiable in advance.

Articles are research reports whose conclusions are of general interest and which are sufficiently rounded to be a substantial advance in understanding. They should not have more than 3,000 words of text (not including figure legends) or more than six display items and should not occupy more than five pages of *Nature*.

Articles start with a heading of 50-80 words written to advertise their content in general terms, to which editors will pay particular attention. The heading does not usually contain numbers, abbreviations or measurements. The introduction to the study is contained in the first two or three paragraphs of the article, which also briefly summarize its results and implications. Articles have fewer than 50 references and may contain a few short subheadings.

Letters are short reports of outstanding novel findings whose implications are general and important enough to be of interest to those outside the field. Letters should have 1,000 or fewer words of text and four or fewer display items. The first paragraph describes, in not more than 150 words and without the use of abbreviations, the background, rationale and chief conclusions of the study for the particular benefit of non-specialist readers. Letters do not have subheadings and should contain fewer than 30 references.

Commentary articles deal with issues in, or arising from, research that are also of interest to readers outside research. Some are commissioned but suggestions can be made to the Commentary Editor in the form of a one-page synopsis.

News and Views articles inform non-specialist readers about new scientific advances, sometimes in the form of a conference report. Most are commissioned but proposals can be made in advance to the News and Views Editor.

Scientific Correspondence is for discussion of topical scientific matters, including those published in *Nature*, and for miscellaneous contributions. Priority is given to letters of fewer than 500 words.

Preparation of manuscripts

All manuscripts should be typed, double-spaced, on one side of the paper only. An original and four copies are required, each accompanied by artwork. If photographs are included, five sets of originals are required; for line drawings, one set of originals and four good-quality photocopies are acceptable. Reference lists, figure legends and tables should all be on separate sheets, all of which should be double-spaced and numbered. Three copies of relevant manuscripts in press or submitted for publication elsewhere should be included with each copy of a submitted manuscript, and clearly marked as such. Five copies of revised and resubmitted manuscripts, labelled with their manuscript numbers are required, together with five copies of a letter detailing the changes made.

Titles are brief and simple. Active verbs, numerical values, abbreviations and punctuation are to be avoided. Titles should contain one or two key words for indexing purposes.

Artwork should be marked individually and clearly with the author's name and, when known, the manuscript number. Ideally, no figure should be larger than 28 by 22 cm. Figures with several parts are to be avoided and are permitted only if the parts are closely related, either experimentally or logically. Unlettered originals of photographs should be provided. Suggestions for cover illustrations, with captions and labelled with the manuscript number, are welcome. Original artwork is returned when a manuscript cannot be published.

Protein/nucleotide sequences should ideally be in the three-letter and not the single-letter code for amino acids. One column width of *Nature* can accommodate 20 amino acids or 60 base pairs.

Colour artwork. A charge of £500 per page is made as a contribution towards the cost of reproducing colour figures. Inability to pay these costs will not prevent publication of essential colour figures if the circumstances are explained. Proofs of colour artwork may be sent to contributors under separate cover from their galley proofs.

Figure legends should not exceed 300 words and ideally should be shorter. The figure is described first, then, briefly, the method. Reference to a method published elsewhere is preferable to a full description. Methods are not described in the text.

References are numbered sequentially as they appear in the text, followed by those in tables and finally by those in figure legends. Only papers published or in the press are numbered and included in the reference list. All other forms of reference should be cited in the text as a personal communication, manuscript submitted or in preparation. Text is not included in reference lists. References are abbreviated according to the *World List of Scientific Periodicals* (Butterworths, London, 1963-65). The first and last page numbers are included; reference to books should include publisher, place and date.

Abbreviations, symbols, units and Greek letters should be identified the first time they are used. Acronyms should be avoided whenever possible and, if used, defined. Footnotes are not used in the text.

Acknowledgements are brief and appear after the reference list; grant and contribution numbers are not allowed.

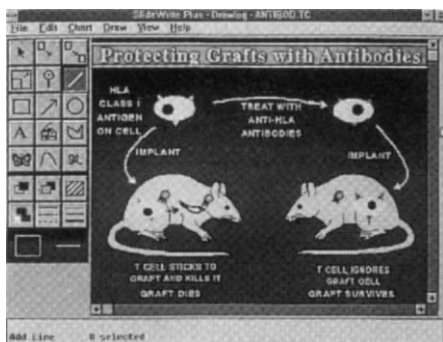
Supplementary information is material relevant to Articles or Letters which cannot, for lack of space, be published in full, but which is available from *Nature* on request.

Submission. Manuscripts can be sent to the Editor at 4 Little Essex Street, London WC2R 3LF, UK or at 1234 National Press Building, Washington, DC 20045, USA. Manuscripts or proofs sent by air courier to London should be declared as 'manuscripts' and 'value \$5' to prevent the imposition of import duty and value-added tax.

Technical toolkits

Making a splash this week — a technical presentation graphics package, a plasmid drawing program, a troubleshooting program for HPLC, molecular modelling software and data acquisition products.

ADVANCED Graphics Software announces the availability of SlideWrite Plus for Windows, a **technical presentation graphics software package** for use with Microsoft Windows 3.0 (*Reader Service No. 100*).



SlideWrite Plus for Windows.

The program combines drawing and text capabilities with scientific graphing, curve fitting, data transformation and statistical analysis capabilities. It has a WYSIWYG user interface and pull-down menus, and permits onscreen drawing, sizing and text formatting. Scalable Nimbus-Q fonts (including Greek and mathematical symbols) and vector-format clip-art ensure that the integrity of body text and drawings is maintained, regardless of the size of the output. SlideWrite Plus charts can be moved easily via the Windows clipboard for inclusion in WordPerfect, Word, Pagemaker or Ventura Publisher for Windows documents. The program will accept data from Lotus.WK1, .WKS and .WK3 files, ASCII files, and from any application using dynamic data exchange. SlideWrite Plus for Windows supports most output devices. A network version is available that supports Novell, Ethernet and other DOS 3.1 or higher compatible networks. The package comes with over 400 logically grouped, vector-format illustrations of animals, arrows, buildings, chemical symbols, computers, flow charts, icons, plants, scientific apparatus and symbols, signs, etcetera. Additional clip-art packages, called Figure Packs, are available (biological cells, chemistry, drugs, human anatomy, maps and a general scientific pack with chemical and laboratory symbols). SlideWrite Plus for Windows costs US\$445; optional Figure Packs cost US\$99 each.

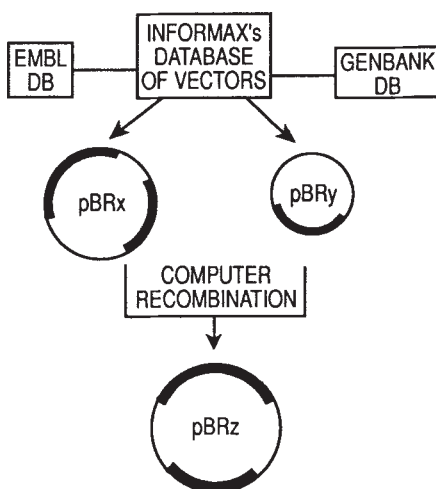
Last month, STN, the Scientific and Technical Information Network, added **21 new databases** (*Reader Service No. 101*). The new listing includes: LIFESCI (biosciences), CBND (chemical business news), NATURE · VOL 361 · 18 FEBRUARY 1993

CheminformRX (chemical reactions), ELCOM (electronics and communication), AQUASCI, OCEAN (aquatic sciences, marine biology and physical oceanography), HEALSAFE (health and safety), COMPUAB (computer and information systems), CONFSCI (conference papers), ISMEC (mechanical engineering), PNI (pharmaceutical news), SOLIDSTATE (superconductivity), DETHERM (thermodynamics), HSDB (toxicology) and POLLUAB (pollution).

Responding to feedback from users, Shandon Scientific has released a new version of **control software for its tissue processor**, the Hypercenter XP (*Reader Service No. 102*). Control functions now include a one-button drain command, giving automatic wax draining at the end of a processing cycle, higher flush reagent temperatures and a simplified routine status display that removes information of less immediate relevance to the operator. Shandon says that the new drain command, for example, should give staff the confidence to leave specimens in the processor at the end of the cycle, without any danger of de-waxing.

■ Molecular biology programs

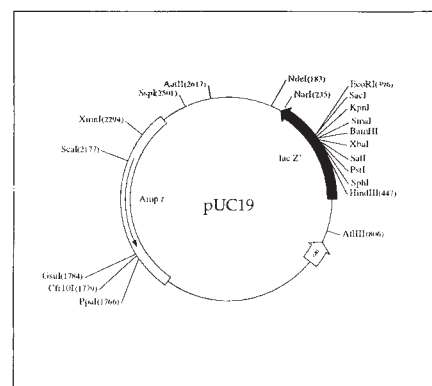
Vector PC is a software system from InforMax that should be of interest to researchers involved in **cloning vector and recombinant DNA research and development** (*Reader Service No. 103*). Vector PC includes a database of 270 well-characterized cloning vectors and other genetic molecules, and tools for genetic engineering design. The program can be used to



Vector PC — a computerized solution for genetic engineering.

construct new recombinant molecules using standard genetic engineering methodologies and to enter them into the database. It can also be used for creating exact nucleotide sequences of new molecules after multistep recombination and fragments termini treatment; for creating and representing graphically functional and restriction maps of new recombinant molecules; and for managing the database of vectors and recombinant molecules. Users can import from other databases (EMBL, GenBank and Entrez). Vector PC runs on IBM PCs or compatible computers and costs US\$995.

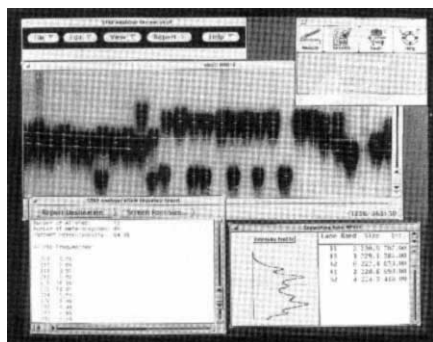
Version 1.2 of Plasmid Artist, a **plasmid drawing program** for Apple Macintosh computers, is now available from Clontech Laboratories (*Reader Service No. 104*). Plasmid Artist allows the user to draw linear or circular maps quickly and easily. Users can create restriction maps from imported



Plasmid Artist can be used for drawing plasmid maps fit for publication purposes.

sequences and the program's menu of over 80 restriction enzymes. The program should also be useful for cloning fragments into vectors. Maps are scaled automatically and can be edited and resized. Fragments can be copied and pasted from one map to another, and maps can be exported into, for example, other drawing or word processing programs. The program will run on any Macintosh computer with at least 1 MB of RAM. A hard disk and PostScript printer are recommended. System 6.05 or higher is required.

Millipore announces the availability of a dedicated software package for **calculating the size of short, repeated DNA sequences** used in genetic mapping (*Reader Service No. 105*). These repeated sequences (known as CA repeats, short tandem repeat polymorphisms (STRPs), microsatellites or sim-



For increased genotyping throughput — see Millipore's software solution.

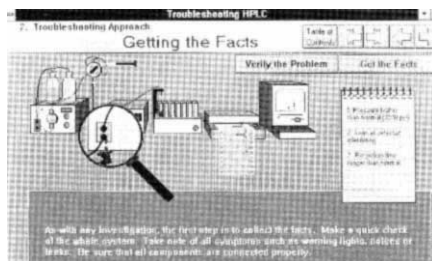
ple sequence repeats) are an increasingly popular alternative to restriction fragment length polymorphisms. Short tandem repeat polymorphisms are highly polymorphic and small amounts can be amplified by the polymerase chain reaction. One drawback to using STRPs has been the need to score manually alleles to determine their size in base pairs. Scoring alleles by eye can be time-consuming and inherently hard to reproduce. The new software package, called Bio Image Short Tandem Repeat Polymorphism software, eliminates this bottleneck, says Millipore, by quickly and objectively scoring alleles. At present, a researcher should be able to calculate manually about 1,000 to 2,000 genotypes per week (that is, score all alleles); Bio Image is intended to increase throughput by about five times to 7,500 genotypes per week. The software also provides supporting data and calculates the allele frequencies for a genetic marker. These data can then be exported in formats compatible with genetic linkage and mapping software. The software runs on all turnkey Bio Image gel/film reader systems, in addition to standalone computers from Sun Microsystems.

■ Chromatography corner

Galactic Industries has improved the capabilities of GRAMS/386 (a Microsoft Windows-based application) with the addition of a **method editor for chromatography applications** (Reader Service No. 106). The GRAMS method editor is menu driven and provides, Galactic says, a 'logical' organization of chromatographic functions. The program can work with a variety of chromatography data from liquid chromatography, gas chromatography (GC), HPLC, size-exclusion chromatography/GPC, diode array/HPLC, GC/mass spectrometry and GC/Fourier transform-infrared spectroscopy. The graphics capabilities found in the GRAMS method editor provide the chromatographer with powerful data visualization screens for displaying information such as calibration curves or timed events. Reports and printouts are generated from within the GRAMS' environment. Calibration reports can be output to hard copy and method files can be printed as they appear in the method editor. Method file printouts are

organized sectionally by calibration, peak picking, run control, etcetera. The GRAMS method editor allows changes and updates to method files. Custom postprocessing routines can be created to supplement the standard functions provided. GRAMS/386 runs under Microsoft Windows versions 3.0 and higher on 386 and 486 PCs and compatible computers. A maths co-processor and VGA display are required.

Troubleshooting HPLC is the new Windows-based **diagnostic and training software program for HPLC** that is available from Autoscribe (Reader Service No. 107). This program features an expert system that is designed to help the user locate the source of a problem in a HPLC system and assist in correction of the problem. After asking the user a series of questions, the responses are recorded and a list of possible causes and corrective actions is presented. A printed report is also available, if required. In addition to its use as an everyday working tool, the program provides computer-based interactive training on HPLC troubleshooting. Animated graphics illustrate pressure problems, blockages, leaks, component fail-



HPLC troubleshooting program.

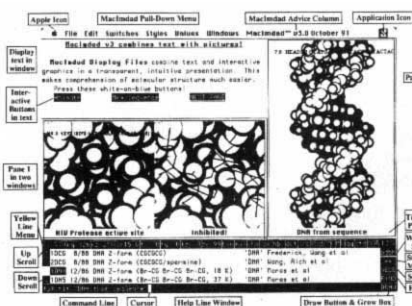
ures, baseline drift, peak distortions and data system malfunctions. Reviews, onscreen quizzes and appendices that can be printed complete the package. **Troubleshooting HPLC** is the fourth in a series of five computer-based programs developed by Dennis Saunders of Unocal Science and Technology Division, Brea, California.

■ Making models

Last month, BioCAD began shipping Version 1.1 of the **Catalyst drug discovery software system**, which is designed for use by medicinal chemists (Reader Service No. 108). Version 1.1 provides chemists and molecular modellers with access to Catalyst from Macintosh computers and PCs that are connected to Silicon Graphics' computers on an Ethernet network. Catalyst is designed to help speed drug discovery by automating the steps in 'hypothesis-based drug design' — the essence of which is to uncover the relationship between chemical structure and biological activity and to then use that information to discover active compounds. With Catalyst 1.1, chemists enter the two-dimensional (2-D) molecular structures and biological activity data of leading molecules using their Macintosh or PC.

They can draw the structures using Catalyst, cut and paste existing ChemDraw structures, or import MDL MOL files. The program generates flexible three-dimensional (3-D) conformational models and then extracts a hypothesis, which is displayed as abstract chemical features positioned in 3-D space that correlate with biological activity. Chemists view molecules automatically superimposed against their hypotheses and design new molecules in order to improve the fit. Catalyst estimates the activity of the new compounds before they are synthesized, allowing medicinal chemists to focus their laboratory efforts on the most promising candidates. An existing database of 2-D chemical structures can be converted into a Catalyst database of 3-D conformational models. The Catalyst database can be searched with the chemist's hypothesis to find novel structural motifs that may lead to new biologically active molecules. Catalyst can be accessed from any Apple Macintosh Quadra with a 16- or 21-inch monitor. Installed Macintosh IIci computers need to be upgraded with high-resolution (1,024 × 768) graphics and a Radius Rocket accelerator board. PCs require the Number Nine GXi graphics accelerator card as a high-resolution monitor.

Molecular Applications Group is now shipping MacImdad Version 3.06, the latest version of its program that provides **access to over a thousand protein and nucleic acid structures** determined by X ray and NMR (Reader Service No. 109). The program provides rapid access to drawings of any molecule in the Brookhaven Protein Data Bank (there are more than 800 to date). Users can create any organic molecule by specifying a simple chemical formula, modify the structure by editing the chemical formula and nonstandard geometry, and output the coordinates for use in molecular mechanics, ligand fitting, drug design, etcetera. In addition, users can construct standard DNA double helices of arbitrary sequence in either A or B form, which should facilitate the visualization of possible protein recognition sites. MAG says that users can model peptides of any sequence and conformation and manipulate them as desired. A full range of drawing styles is included. The animation feature can be used



Imdad — MAG's interactive display and design software is now available for Macs.

to display a sequence of frames to show rotating images built up from a combination of stick bonds, space-filling atoms and dot surfaces. This, MAG says, provides a vivid impression of molecular shape and bulk in three dimensions. MacImdad runs on Macintosh computers from the LC to the Quadra. Imdad is also available for all Silicon Graphics' workstations. SGImdad provides users with the same functionality as MacImdad but, more importantly, provides easy access to MAG's range of high-end simulation and modelling software.

Molecular Arts has announced the availability of chemVISION for Windows Version 2.0, a PC software tool for three-dimensional (3-D) **molecular visualization** (*Reader Service No. 110*). The program offers 3-D rendering and chemical query capabilities and is designed specifically for academics, engineers, researchers, scientists and students in the chemistry field. With chemVISION, chemists can tap into 3-D chemical structure databases, display and render molecular structures in three dimensions, rotate and scale molecules in 3-D space, examine and make queries of their molecule, illustrate their research documentation with 3-D graphics and produce output of a presentation quality. chemVISION utilizes the Microsoft Windows graphical user interface. This version features several



Buckminsterfullerene 'Bucky Balls' (C₆₀) shown as a ball-and-stick model with smooth shaded spheres.

new or upgraded capabilities: chemREAD-3D for the import/export of 3-D molecular structure files and for translating data into different formats; SELECT-3D, a molecular substructure identifier and selector; chemGRFX for handling vector- and raster-based molecular graphic images; and VIS-AGE-3D, a set of molecular rendering tools. A single-user licence for chemVISION costs US\$299.

■ Data capture

Keithley Instruments announces the availability of Iotech's DaqBook/100 — a **data acquisition add-on product** that supplies high-speed, multifunction I/O capability to notebook PCs for portable test applications



High-speed data acquisition product for notebook PCs.

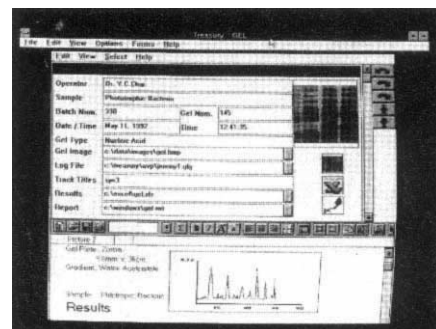
(*Reader Service No. 111*). It is also designed to provide a cost-effective alternative to plug-in boards for conventional desktop PC-based data acquisition applications. The DaqBook/100 can transfer data bidirectionally across the PC's parallel port at up to 170 kB s⁻¹, permitting real-time acquired data to be stored in the PC's memory and on its hard drive. It offers a variety of data acquisition capabilities, including a 100-kHz 12-bit, 16-channel A/D converter with a programmable gain input amplifier; two 12-bit D/A converters; five 16-bit counter-timers; 16 digital inputs that can be sampled at up to 100 kHz; and 24 general-purpose digital I/O lines. Users can expand channel capacity to 256 analog input channels via optional expansion modules. The DaqBook/100 can be powered via several sources, including a standard 12-V battery, an included a.c. adapter or an optional rechargeable nickel-cadmium battery module. This makes it suitable for portable and remote data acquisition applications.

Cambridge Electronic Design has released Spike2 for Apple Macintosh computers, **data acquisition and analysis package** for the research laboratory (*Reader Service No. 112*). The Spike2 package for the Macintosh allows the recording of 18 channels for waveform and digital data with scrolling chart-recorder style displays and continuous sampling to hard disk at rates of up to 80 kHz. The analysis section allows for spike train processing, including peristimulus time histograms, correlations, raster displays and frequency plots, spike shape capture and waveform averaging with respect to events. The program requires a Cambridge Electronic Design 1401 *plus* laboratory interface, which contains its own 32-bit processor and up to 16 MB of memory. This set up allows many data analysis tasks to be performed online, releasing processor time for the Macintosh.

Axon Instruments announces the release of its AxoTape 2.0 software — a data acquisition program that provides a **multichannel recording and monitoring system for real-time analog signals** (*Reader Service No. 113*). Large amounts of data can be displayed, digitized and recorded without interruption, or sweeps of data can be acquired in an oscilloscope-like manner. AxoTape 2.0 has been designed, the company says, as a low-cost alternative to digital oscilloscopes, analog storage oscilloscopes, VCRs, chart recorders and tape recorders for a wide variety of recording situations. With AxoTape 2.0, acquisition of fixed-length data sweeps can be triggered by an incoming signal, external TTL pulses or by pressing a key. The program can also autotrigger if a trigger is not detected. It can update a cumulative average of sweeps and/or maintain previous sweeps of data onscreen for a 'persistence' display. During recording, the operator can tag specific time points via the keyboard or with TTL pulses and instantly jump these time points when reviewing the data record. These tags can also be included in data plots. Data files can be played back for review, cursor measurements or for output to printers, plotters and ASCII files. The data files can be read directly by the company's pCLAMP for electrophysiological single-channel analysis or by other spike analysis software packages. A programmers' support library of data conversion routines is also provided. AxoTape 2.0 runs on PC/AT and compatible computers. Up to sixteen multiplexed analog input channels can be acquired. With Axon Instruments' Digidata 1200 data acquisition interface, AxoTape 2.0 can acquire data at a 12-bit resolution over a ± 10 -V range and at rates of up to 333 kHz. The list price for AxoTape 2.0 is US\$1,000.

■ Imaging software

Running under Windows 3.1, the new **image database** called Treasury from Synoptics is designed to manage the burgeoning quantity of scientific image data, including photographs, pictures, drawings, reports and records (*Reader Service No. 114*). For flexibility, Treasury incorporates a graphical tool for designing the way data and images are entered into the database and



A typical Treasury Viewer with a combination of Excel spreadsheet, image and data.

for creating the 'viewers' that display the image and its associated data. If required, Synoptics says that different viewers can be linked together in a hierarchy, providing a means of moving from one image to another. An initial viewer might, for example, show the original histological sample at low magnification; a second and third viewers, images at high magnification of two particular regions of interest with associated text descriptions; and a final viewer, a three-dimensional reconstruction of the sample. Treasury also includes a means of browsing and searching for selected images. As many users may already have images filed, Treasury allows access to existing image files by using simple reference keys. These files can be off-line if required, for example, in a cabinet of floppy disks.

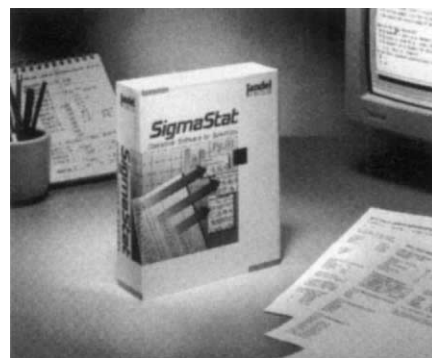
Confocal Technologies' range of scientific **imaging software**, Vision Science, is a comprehensive set of products comprising general purpose utilities and special quantitative imaging applications running under Windows 3.0 on the PC platform (*Reader Service No. 115*). For establishments that already have the capability to digitize images, Cyclops is a useful resource for image processing and analysis with facilities to import digital images of many formats. Running on single screen and using only the computer's RAM and own VGA/Supra VGA graphics, the UK£395 (UK£295 academic price) Cyclops is intended to be a low-cost, entry-level imaging software product. Fenestra Vision offers high-quality video digitization and is supported by SuperGrab, a resource for flexible image acquisition formats, averaging, timed series, animations and high-speed image sub-region acquisition. This is combined with image processing functions with fully calibrated geometrical and densitometrical analysis, and image publishing resources. For three-dimensional (3-D) reconstruction and rendering, Microvoxel, with OS2/2.0-based software and specialized hardware for the PC, provides surface and volume rendering from sections in addition to 3-D image processing and analysis. Digital

Stereology and Comet are designed for users with specific imaging objectives — to obtain unbiased 3-D measurements from images of sections and to quantify the Comet Assay for DNA damage, respectively.

■ Numbers game

Macsyma is now shipping a new version of its Macsyma **mathematical software** program for Windows 3.1, a program that resulted from research at the Massachusetts Institute of Technology (*Reader Service No. 116*). The updated version includes enhanced scientific graphics capability and near-typeset quality math display. Macsyma's scientific graphics have been improved: users can make parametric plots of three-dimensional objects with control of colour, lighting, rotation, translation, zoom and other features. The company says that the program's graphics viewer enables users to peer inside plotted objects with realistic perspective. The graphical user interface allows near-typeset quality display of mathematical expressions and online help that includes over 1,500 hypertext command descriptions, 600 executable demonstrations that can be launched from the hypertext, 400 command templates, and mathematical topic menus that enable the user to select from over 50 topic areas. The updated version also includes significant mathematical advances in areas such as trigonometric simplification, integration, solution of ordinary differential equations and linear algebra.

SigmaStat, a **statistical package** designed for scientists, is the latest addition to Jandel Scientific's line of software tools (*Reader Service No. 117*). The US\$395 SigmaStat program can be used to perform *t*-tests, analysis of variance, correlation, rates and proportions, nonparametric methods, linear and nonlinear regressions, power and sample size. Jandel Scientific says that SigmaStat's procedures are highly automated and require very little statistical knowledge from the user. The program's 'Advise' command helps the user select the most appropriate statistical procedure and then automatically performs the test. SigmaStat also checks that the users' data fit the assumptions underlying the selected statistical methods, warns the user if they are violated and automatically suggests and computes more appropriate alternative tests. The program also handles missing and unbalanced data using a general linear model. SigmaStat calculates a range of descriptive statistics, including mean, standard deviation, standard error of the mean, range, maximum, minimum, normality, median, percentiles, sum of squares, skewness and kurtosis. In addition to these statistical procedures, the program provides several different options for each of its tests. The user can, for example, specify different data formats, including raw data, blocks of data, indexed data and sample size, mean and



Statistical software for the scientist.

standard deviation, or standard error of the mean. The user can decide whether or not to test for normality or equal variance and acceptable *P*-value levels for these tests. The user can choose from a range of regression diagnostic options, specify a maximum order for polynomial regressions, and set the alpha for power calculations.

PS-Plot is the new **technical plotting and data processing software** package, which is completely in graphics mode for DOS, from Polysoft (*Reader Service No. 118*). The program uses a new graphical user interface with most of the standard Microsoft Windows interface power: drop-down menus, multiple windows, mouse support, page layout with WYSIWYG graphics display, dialogue boxes, online help and status bar, and a comprehensive help system. PS-Plot has a comprehensive data sheet that performs statistical analysis (*t*-test, *f*-test, ANOVA, regression), data transformation, digital signal processing (filter, window, smooth), nonlinear parameter fitting, and model development. Data can be exchanged to and from most spreadsheet file formats, including Lotus 1-2-3, dBase, Quattro Pro, DIF, Microsoft Excel, ASCII and SYLK. A plot editing toolbox is provided to facilitate the design and modification of graphs. Various plot types are offered, including two- (2-D) and three-dimensional (3-D) scatter plots, curve (line) graphs, surface (mesh) graphs, 2-D and 3-D vertical/horizontal bar charts, 2-D and 3-D histograms, box plots, step plots, and error bars. Multiple coordinate systems, axes and plots can be placed on one page. The program produces publication-quality graphics output to dot-matrix printers, laser printers, plotters and slide makers. PC-Plot also produces colour EPS, CGM, TIFF, BMP, PCX and HPGL files. An IBM PC, XT, PS/2, AT or compatible computer is required. PS-Plot costs US\$449. □

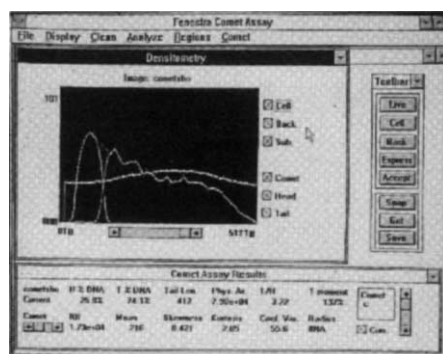


Image analysis screen from the Comet Assay offering quantitative analysis of cells prepared for the assay for DNA damage and viewed in an epifluorescence microscope.

These notes are compiled by Diane Gershon from information provided by the manufacturers. For more information, fill in the reader service card bound inside the journal. Prices quoted are sometimes nominal and apply only within the country indicated.